Aggregating Distortions in Networks with Multi-Product Firms*

Yasutaka Koike
UNC Chapel Hill

Antonio Martner
Central Bank of Chile

June, 2026

Abstract

We investigate the role of multiproduct firms in shaping resource misallocation and its impact on aggregate total factor productivity (TFP) growth. Using administrative data on product transactions between all formal Chilean firms, we provide evidence that demand shocks to one product affect the production of other products within the same firm, suggesting that firms engage in joint production. We develop a framework to measure resource misallocation in production networks with joint production, deriving sufficient statistics to quantify these effects. Applying the framework to a period of negative TFP growth in Chile, we find that the standard single-product benchmark attributes a 14.1% decline to allocative efficiency. Accounting for joint production reduces this measured decline to 10.1%. Ignoring joint production therefore overestimates the decline in allocative efficiency by 40% relative to the joint-production estimate.

JEL Codes: D24, D57, E23, L11, L25, O47.

*We are deeply indebted to David Baqaee, Ariel Burstein, Hugo Hopenhayn, Michael Rubens, and John Asker for excellent guidance. Jonathan Vogel, Oleg Itskhoki, Andy Atkeson, Lee Ohanian, Saki Bigio, Pablo Fajgelbaum, Fernando Alvarez, Javier Cravino, Colin Hottman, Aaron Flaaen, Xiang Ding, Yuhei Miyauchi, Federico Huneeus, Kosuke Aoki, Scott Orr, Swapnika Rachapalli, Ken Kikkawa, and Dan Cao provided valuable comments. We thank seminar participants at the ASSA Annual Meeting, Society for Economic Dynamics Winter Meeting, the Third Global Economic Network Conference, the GSE-OSIPP-ISER Joint Conference, SEA 94th Annual Meeting, the Midwest Macroeconomics Meeting, Penn State University, Keio University, Waseda University, Hitotsubashi University, California State University Long Beach, UBC Sauder, the Federal Reserve Board, 31st Conference on Computing in Economics and Finance and the Central Bank of Chile. The views expressed are those of the authors and do not necessarily represent the views of the Central Bank of Chile or its board members. Authors emails: yasu@unc.edu, amartner@bcentral.cl

1 Introduction

Resource misallocation across heterogeneous producers has been recognized as a driver of aggregate total factor productivity (TFP) growth differences across countries and over time. To quantify the extent of misallocation, recent literature has made extensive use of granular firm-level data.

However, despite its emphasis on micro data, this literature typically ignores the fact that most firms sell multiple products. For example, 75% of formal firms in Chile report selling more than one product, and these firms collectively account for 99% of all firm-to-firm transaction value in Chile using product-level VAT data for 2018. The ubiquity of multi-product firms introduces new challenges to understanding resource allocation. Specifically, researchers must consider how the allocation of resources across products within firms affects allocative efficiency and aggregate TFP growth. Measuring resource allocation within firms requires in principle determining how to map inputs to specific outputs.

The literature on multi-product firms typically assumes product line independence (Klette and Kortum (2004); Bernard et al. (2011); De Loecker et al. (2016); Hottman et al. (2016); Mayer et al. (2021)). If firms are collections of independent products, the challenge is reduced to a measurement problem, and existing theories for single-product firms can be applied by treating different products as if they are separate firms. However, firms often simultaneously produce multiple outputs using shared inputs, making it impossible to assign inputs to specific outputs. Consider an oil refinery that produces diesel and gasoline concurrently: the inputs—crude oil, labor, and capital—are used to produce both outputs and cannot be accounted for separately.

How do multi-product firms with non-separable production technologies affect the measurement of the extent of resource misallocation? We model firms' technology via joint non-separable production functions that map bundles of inputs into bundles of multiple outputs. This approach eliminates the need to define individual product-level production functions. The joint production function describes the firm's flexibility in adjusting its product mix, which then determines the extent of resource allocation within a firm.

We generalize previous work (e.g, Baqaee and Farhi (2020)) to accommodate multi-product firms with joint production technologies. We provide sufficient statistics to measure changes in allocative efficiency using an ex-post strategy. Our framework is general enough to accommodate firm-to-firm linkages. We provide empirical evidence for joint production and implement our framework using product-level firm-to-firm transaction data from Chile. We show that the extent of resource misallocation is overstated if we abstract from joint production, as is standard practice in the literature. While our primary focus is ex-post analysis, we also develop a complementary parametric framework to understand ex-ante counterfactuals, such as the gains from eliminating distortions, taking into account joint production technologies.

We first describe the theoretical contributions of the paper, then turn to empirical validation and application. In our model, products within firms can be under- or overproduced because they face different wedges (e.g., markups). Loosely speaking, products with relatively high wedges are underproduced relative to an efficient economy. However, since firms are embedded in supply chains, the relevant wedges that affect resource allocation and, hence, aggregate TFP growth are not just the firm's own wedges, but the entire chain of cumulative wedges that travel from final demand upstream to where production occurs.

The impact of changes in these cumulative wedges on resource misallocation depends crucially on how easily firms can adjust their product mix. Consider again the oil refinery example. If the oil refinery raises the markup on gasoline, thereby lowering its demand, it cannot redirect production resources toward diesel because the production technology yields gasoline and diesel in nearly fixed proportions from crude oil. This technological constraint limits the firm's ability to reallocate resources in response to demand changes and hence limits the extent of misallocation within the firm. Therefore, joint production technology can attenuate the extent of resource misallocation and its contribution to aggregate TFP growth.

To quantify the extent of resource misallocation in the presence of joint production, we develop a sufficient statistics approach. Our approach relies on observed changes in product-level prices within firms. Theoretically, these price movements, net of markups, trace out the production possibility frontier, whose slope captures each firm's technological constraints when adjusting their product mix. When firms have flexibility to adjust their product mix, changes in prices net of markups will be small as firms can easily substitute production between products.

The covariance between relative price changes and cumulative wedges at the product level captures the attenuation of resource misallocation due to joint production technology. Intuitively, if prices rise for products with high (cumulative) wedges, then the scope for reallocation is limited. Rather than directly estimating the production possibility frontier, our approach infers its shape from observed price changes, providing a way to quantify misallocation without imposing parametric assumptions about firms' production technologies.

The growth-accounting formula we develop generalizes previous approaches: it collapses to the Baqaee and Farhi (2020) growth accounting result under the single-product firm assumption and to the Hulten (1978) benchmark under perfect competition.

To implement our framework, we use administrative firm-to-firm transactions data from the Chilean Internal Revenue Service. The dataset contains product-level prices, firm-product input-output linkages, and balance sheet variables — enough to construct our sufficient statistics.

We first provide evidence of joint production in the data. We examine whether the standard assumption in the literature — that firms operate as independent single-product lines — holds. We find that demand shocks to one product significantly affect the production of other products within the same firm, consistent with joint production. Specifically, we find that a negative demand shock to a firm's main product reduces the quantities and raises the prices of other products, as predicted by our framework.

Implementing our framework requires product-level own wedges. We therefore construct product-level markups by adapting the production approach of Dhyne et al. (2022), which builds on the production-set framework of Diewert (1971), to our joint-production environment. The advantage of our data is that we observe product quantities, not only revenues, which allows us to measure heterogeneous product-level own wedges within firms. Product-level markups are used to construct the own-wedge component of cumulative wedges; the main sufficient statistic combines cumulative wedges with observed within-firm product price changes.

Having provided evidence of joint production and constructed the required own-wedge component, we implement our sufficient statistics to conduct a growth accounting exercise for Chile during a period in which aggregate TFP declined by 8.5%. A standard approach that abstracts from joint production attributes a 14.1% decline to allocative efficiency from 2016 to 2022. Accounting for joint production reduces this measured decline to 10.1%. Ignoring joint production therefore provides a misleading assessment and overestimates the decline in allocative efficiency by 40%. This finding reflects constraints faced by multiproduct firms when adjusting their product mix. These constraints limit the scope for product-level resource reallocation within firms, and this affects allocative efficiency and aggregate TFP growth.

We complement the ex-post analysis with a parametric framework for ex-ante counterfactuals. We derive closed-form expressions for the distance to the Pareto-efficient frontier under joint production. Consistent with our earlier finding that joint production constrains reallocation, we find that assuming separable production technologies overestimates the extent of resource misallocation caused by a given set of wedges.

Related Literature

Our paper contributes to the literature on misallocation and aggregate productivity. Following Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), a large literature studies how wedges across producers affect aggregate TFP. Recent work shows that input-output linkages are central for measuring these losses, because distortions propagate through production networks (Baqaee and Farhi, 2020; Bigio and La'O, 2020).

We build on this approach by allowing firms to produce multiple outputs with joint production technologies. This matters because misallocation depends not only on wedge dispersion across firms but also on within-firm product-mix adjustment: when products face different cumulative wedges, the extent to which firms can reallocate production across products determines how distortions aggregate. Our sufficient-statistics formula therefore generalizes existing growth-accounting results (Solow, 1957; Hulten, 1978; Basu and Fernald, 2002; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2020; Baqaee et al., 2023) to production networks with multiproduct firms and joint production.

The paper also relates to the literature on multiproduct firms. This literature studies product scope, product switching, demand, and markups within firms (Bernard et al., 2010, 2011; Mayer et al., 2014; De Loecker et al., 2016; Hottman et al., 2016; Mayer et al., 2021; Wang and Yang, 2023). A common assumption in this work is that firms can be treated as collections of separable product lines. In contrast, we focus on the aggregate implications of nonseparable production across products. In our framework, the firm is not simply a portfolio of independent products: common inputs and joint production technologies constrain product-mix adjustment, altering how product-level wedges translate into aggregate misallocation.

Our approach connects this multiproduct-firm literature to the literature on joint production and multi-output technologies (Powell and Gruen, 1968; Diewert, 1971; Lau, 1972; Hall, 1973, 1988; Boehm and Oberfield, 2023; Carrillo et al., 2023; Ding, 2023). Recent work has developed methods to estimate production functions for multiproduct firms with joint production (Dhyne et al., 2017, 2022; Valmari, 2023; Cairncross and Morrow, 2023). We build on these estimation methods for measurement, but our main contribution is different: we derive aggregate sufficient statistics showing how joint production changes the mapping from observed wedges to allocative efficiency in a production network. Empirically, we use comprehensive Chilean product-level transaction data to implement the framework.

Unlike prior work that imputes network linkages from industry-level input-output tables (for example, Baqaee and Farhi (2020) combine US Compustat data with industry-level IO tables) we directly observe product-level firm-to-firm transactions. This allows us to construct firm-product network linkages, measure product-specific prices and quantities, and estimate cumulative wedges without relying on industry aggregates.[1]

Our empirical evidence on joint production is related to the literature on shock transmission. Much of the firm-to-firm network literature studies how supply shocks propagate downstream across firms (Boehm et al., 2019; Carvalho et al., 2020; Fujiy et al., 2022; Bai et al., 2024). We instead use external demand shocks to test whether shocks to one product affect the production and pricing of other products within the same firm. This connects our evidence to studies of within-firm spillovers from demand shocks (Giroud and Mueller, 2019; Almunia et al., 2021; Ding, 2023), but our focus is on using these spillovers to discipline joint production and quantify its aggregate implications.

2 A Theory of Distortions Aggregation with Multiproduct Firms

We develop a theoretical framework to analyze resource misallocation in production networks with multiproduct firms that use joint production technologies. Given that firms use shared inputs to produce multiple outputs simultaneously, even with suitable data, it is impossible to assign

[1] Using the same Chilean transaction data, Burstein et al. (2024) study misallocation from buyer-specific prices for the same product; our paper instead focuses on how joint production within multiproduct firms affects the aggregation of distortions. Using Belgian firm-to-firm transaction data, Kikkawa (2022) studies firm-pair-specific markups.

inputs to products separately. We generalize previous frameworks to accommodate multi-product firms within production networks with joint production technologies. We provide sufficient statistics to measure changes in allocative efficiency using an ex-post strategy.

The section proceeds as follows. First, we formalize the concept of joint production technology. To build intuition, we present parametric examples illustrating how joint production affects resource allocation and aggregate TFP. Then, using a parametric model data generation process, we discuss the sufficient statistics needed to measure allocative efficiency. Next, we present our general model, and main proposition, which unpacks allocative efficiency into a single-product term and a multiple-product term.

2.1 Joint Production

We begin by formalizing the concept of joint production, where firms simultaneously use shared inputs to produce different products. To formalize this concept, we follow Hall (1973)'s approach to joint production technology. Let $J(\boldsymbol{q}, \boldsymbol{x})$ be a joint production function, where $\boldsymbol{q}$ is a vector of outputs and $\boldsymbol{x}$ is a vector of inputs. The joint cost function is derived from the firm's cost minimization problem, as follows:

$$C(\boldsymbol{q}, \boldsymbol{p}) \equiv \min_{x \in V(q)} \boldsymbol{p}'\boldsymbol{x},$$

where $V(\boldsymbol{q})$ is the input requirement set, $V(\boldsymbol{q}) = \{\boldsymbol{x} | J(\boldsymbol{q}, \boldsymbol{x}) \geq 0\}$ and $\boldsymbol{p}$ is a vector of input prices. We introduce two assumptions about the shape of the joint production function, which will be used throughout this paper.

Assumption 1 (Constant Return to Scale (CRS))**.** $J(\boldsymbol{q}, \boldsymbol{x}) = 0$ implies $J(\lambda \boldsymbol{q}, \lambda \boldsymbol{x}) = 0$ for any scalar λ. [2]

Unlike a single-output production function, the output is a vector. Note that we do not assume CRS for each single-output production function.

Assumption 2 (Separability between Input and Output Functions)**.** The joint production function can be written as $J(\boldsymbol{q}, \boldsymbol{x}) = -F^Q(\boldsymbol{q}) + F^X(\boldsymbol{x})$, and the joint cost function as $C(\boldsymbol{q}, \boldsymbol{x}) = H(\boldsymbol{q}) \varphi(\boldsymbol{p})$.

Note that this differs from assuming separable production functions, where the output, q, is a single product, not a vector; it degenerates to $F^Q(q) = q$. Example 1 illustrates a joint production function satisfying assumptions 1 and 2:

[2] While we assume CRS, this is not theoretically restrictive. Variable returns to scale can be accommodated through constant returns and producer-specific fixed factors. However, our empirical application to Chile adopts constant returns with respect to observable inputs — labor, capital, and intermediates

Example 1 (Constant Elasticity of Transformation Output Bundle and Constant Elasticity of Substitution Input Bundle (CET-CES))**.**

$$\underbrace{\left(\sum_g q_g^{\frac{\sigma+1}{\sigma}}\right)^{\frac{\sigma}{\sigma+1}}}_{\text{Output bundle}} = A\underbrace{\left(L^{\frac{\theta-1}{\theta}} + K^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}}_{\text{Input Bundle}},$$

The associated cost function is

$$C\left(\boldsymbol{q}, w, r\right) = \frac{1}{A}\left(\sum_g q_g^{\frac{\sigma+1}{\sigma}}\right)^{\frac{\sigma}{\sigma+1}} \left(w^{1-\theta} + r^{1-\theta}\right)^{\frac{1}{1-\theta}},$$

where L and K are the two inputs, w and r are their prices, and $\boldsymbol{q}$ is a vector of outputs.

The input bundle takes a standard CES function with elasticity of substitution θ, and the output is a vector of products rather than a scalar. The parameter σ is called the constant elasticity of transformation; it gives a constant value to the production possibility frontier's curvature of the products within a firm. This example is illustrative as our theoretical framework requires no parametric assumption.

2.2 Parametric Examples of Misallocation with Joint Production

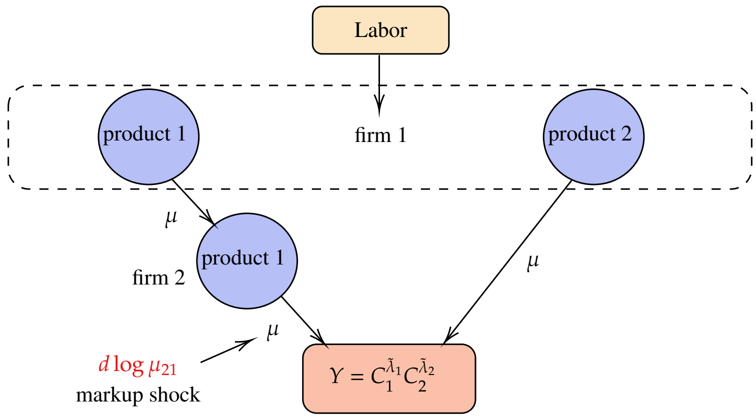
Before presenting the general framework, We provide simplified examples to help gain an intuition about how joint production affects resource allocation and aggregate TFP.[3] Proofs are provided in Appendix F.

2.2.1 An Example without Joint Production

We begin with an example of a production network with multiproduct firms but without joint production. Consider an economy with two firms, as illustrated in Figure 1.

[3] We use markup and wedge interchangeably to refer to any distortion that creates a gap between price and marginal cost. This can include any distortions such as taxes, subsidies, or financial frictions.

Figure 1: A simplified economy with production networks and multiproduct firms



Firm 1 uses labor to produce two differentiated products using labor (L) as unique input, $q_{11} = L_{11}$, $q_{12} = L_{12}$, where $L = L_{11} + L_{12}$. Product 1 is sold to firm 2, while product 2 is sold directly to households. For simplicity, we assume that both products have the same markup, μ. Firm 2 uses product 1 from firm 1 as a production input and produces a different product using a linear technology ($q_{21} = q_{11}$) that sells to households with markup μ. Final consumption goods are aggregated using a Cobb-Douglas function $Y = c_1^{\tilde{\lambda}_1} c_2^{\tilde{\lambda}_2}$, where $c_1 = q_{21}$, $c_2 = q_{12}$. In this simple economy, Y is the real GDP, and aggregated TFP can be defined as $TFP = Y/L$.

In a production network environment, distorted resource allocation arises from both firms' own markups and downstream firms' markups. In this example, product 1 is sold to households with double marginalization: firm 1 charges a markup to firm 2, and firm 2 charges a markup to the household, which inherits firm 1's markup through its marginal cost. As a result, product 1 from firm 1 suffers from a higher distortion than product 2 from firm 1, both relative to a perfect competition setup. To capture this concept in our simplified economy, we introduce the notion of cumulative wedges.

Definition 1 (Cumulative Wedge in a Simplified Economy)**.** In this simplified two-firm economy, we define the cumulative wedge for each product g produced by firm 1, denoted by Γ_{1g}, as the product of markups along the production path from the initial producer to the final consumer. Specifically:

$$\Gamma_{11} = \mu^2 \quad \text{and} \quad \Gamma_{12} = \mu,$$

where μ represents the markup applied by each firm. Here, Γ_{11} captures the cumulative wedge for product 1, which is sold to firm 2 before reaching households, thus incorporating both firm 1's and firm 2's markups (double marginalization). In contrast, Γ_{12} represents the wedge for product 2, which is sold directly from firm 1 to the household with only a single markup.

Now consider a shock that changes firm 2's markup on product 1 ($d \log \mu_{21}$). The first-order response of aggregate TFP to this markup shock can be expressed as:

$$\Delta \log TFP = \tilde{\lambda}_1 \left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1 \right) d \log \mu_{21},$$

where $\bar{\Gamma}_1$ is the weighted harmonic mean of cumulative wedges, defined as:

$$\bar{\Gamma}_1 = \left(\tilde{\lambda}_1 \Gamma_{11}^{-1} + \tilde{\lambda}_2 \Gamma_{12}^{-1} \right)^{-1}.$$

Since $\Gamma_{11} = \mu^2 > \Gamma_{12} = \mu$, we know that $\Gamma_{11} > \bar{\Gamma}_1$, making the term $(\bar{\Gamma}_1/\Gamma_{11} - 1)$ negative. Consequently, an increase in markup ($d \log \mu_{21} > 0$) for product 1, which already faces higher cumulative distortions, reduces aggregate TFP. This occurs because the markup increase further distorts the allocation of resources away from the more distorted product, exacerbating existing misallocation.

Conversely, a decrease in markup ($d \log \mu_{21} < 0$) for product 1 increases aggregate TFP. The reduction in markup allows for increased production of product 1, which was previously underproduced due to higher cumulative distortions. As the relative price of product 1 decreases, households shift consumption away from product 2 toward product 1 (through firm 2's product), leading to improved allocative efficiency and higher aggregate TFP.

2.2.2 An Example with Joint Production

Next, we introduce joint production into our simplified economy. Instead of separable production functions for each product, firm 1 uses a joint production technology to produce both products simultaneously using a constant elasticity of transformation (CET) function:

$$\left(q_{11}^{\frac{\sigma+1}{\sigma}} + q_{12}^{\frac{\sigma+1}{\sigma}} \right)^{\frac{\sigma}{\sigma+1}} = L,$$

where σ represents the elasticity of transformation between the two products.

Consider the same markup shock to product 1 from firm 2 as in the previous case ($d \log \mu_{21}$). By taking a first-order approximation of the change in TFP, we obtain the following response:

$$\Delta \log TFP = \left(1 - \frac{1}{\sigma + 1} \right) \tilde{\lambda}_1 \left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1 \right) d \log \mu_{21}, \tag{1}$$

where $\bar{\Gamma}_1$ is defined as before. This expression reveals how joint production affects the transmission of markup shocks to aggregate TFP.

The magnitude of the TFP response depends critically on the elasticity of transformation σ, which governs how easily firm 1 can adjust its product mix. Since $\Gamma_{11} > \bar{\Gamma}_1$, an increase in the markup ($d \log \mu_{21} > 0$) reduces TFP by distorting the allocation of resources away from the more distorted product 1, while a reduction in the markup ($d \log \mu_{21} < 0$) increases TFP by shifting production toward the more distorted product 1. However, joint production attenuates these TFP responses through the term $(1 - \frac{1}{\sigma+1})$. This attenuation factor reduces the magnitude of the TFP response regardless of the direction of the markup shock, reflecting the technological constraints firms face when adjusting their product mix.

The role of these technological constraints becomes particularly clear when we examine two extreme cases:

$$\Delta \log TFP = \begin{cases} \tilde{\lambda}_1 \left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1 \right) d \log \mu_{21} & \text{if } \sigma \to \infty \\ 0 & \text{if } \sigma \to 0. \end{cases}$$

As σ approaches infinity, the case converges to our previous example without joint production. The firm can freely adjust its product mix in response to the markup shock, allowing for the maximum possible reallocation of resources. The TFP response in this case represents the upper bound of the potential efficiency impact.

Conversely, as σ approaches zero, the production technology becomes Leontief in outputs, meaning the products must be produced in fixed proportions. Consider, for example, an oil refinery that produces both gasoline (product 1) and diesel (product 2). Due to the chemical properties of crude oil and technological constraints of the refining process, the refinery cannot easily adjust the ratio of gasoline to diesel production in response to price changes. In this limit case, even if relative prices change due to markup shocks, the firm cannot adjust its product mix, eliminating any potential gains or losses from resource reallocation.

Joint production thus attenuates the TFP response to markup shocks. Relative to the independent-product-line benchmark ($\sigma \to \infty$), the response is scaled by the factor $\frac{\sigma}{\sigma+1} = 1 - \frac{1}{\sigma+1} < 1$, which shrinks its magnitude whether the markup shock is positive or negative. This suggests that previous studies, which implicitly assume infinite substitutability across products ($\sigma \to \infty$), may overestimate the impact of misallocation — both positive and negative — on aggregate TFP.

2.2.3 Towards a Theory for Measurement

While our previous results help us understand how joint production affects TFP responses, the structural results depend on the elasticity of transformation σ, which is difficult to estimate. We seek to express these results in terms of prices, which are easier to obtain from data.

With joint production, changes in relative prices are associated with changes in production ratios:

$$d \log(p_{11}/p_{12}) = \frac{1}{\sigma} d \log(q_{11}/q_{12}). \tag{2}$$

These price movements effectively trace out the production possibility frontier, whose slope captures each firm's technological constraints when adjusting their product mix.

In our simple example with joint production, let λ_{ij} denote the GDP share of product j of firm

i. The downstream markup change implies that the GDP share of product 1 of firm 1 changes by $d\log\lambda_{11} = -d\log\mu_{21}$. Due to the Cobb-Douglas specification of final demand, the GDP share of product 2 does not change ($d\log\lambda_{12} = 0$). Combining these observations with the relationship between relative prices and quantities gives us[4]:

$$d\log(p_{11}/p_{12}) = -\frac{1}{\sigma+1}d\log\mu_{21}. \tag{3}$$

Using this relationship, we can rewrite the TFP response in equation (1):

$$\Delta\log TFP = \left(1-\frac{1}{\sigma+1}\right)\tilde{\lambda}_1\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)d\log\mu_{21}. \tag{4}$$

This can be further decomposed into two terms:

$$\Delta\log TFP = \underbrace{\tilde{\lambda}_1\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)d\log\mu_{21}}_{\text{Single-Product Term}} + \underbrace{\tilde{\lambda}_1 d\log(p_{11}/p_{12})\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)}_{\text{Multi-Product Term}}.$$

These terms can be expressed in terms of observable variables. The single-product term becomes:

$$\tilde{\lambda}_1\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)d\log\mu_{21} = -d\log\Lambda - \tilde{\lambda}_1 d\log\mu_{21}.$$

where Λ is the labor share. The multi-product term can be written as a covariance between price changes and cumulative wedges:

$$\tilde{\lambda}_1 d\log(p_{11}/p_{12})\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right) = \mathrm{Cov}_{s_1}\left(d\log p_{(1,\cdot)}, \frac{\bar{\Gamma}_1}{\Gamma_{(1,\cdot)}}\right).$$

The single-product term captures the resource misallocation effects that would exist even in an economy without joint production. When the initial equilibrium is inefficient, products with high markups are underproduced. A decline in factor shares indicates resources shifting toward these high-markup activities, but we must adjust for mechanical changes in factor shares caused directly by markup changes.

The multi-product term captures how joint production affects firms' ability to reallocate across products. Instead of estimating the firm's production technology parameters directly, we rely on observed changes in product-level prices within the firm. Intuitively, if prices for certain products rise within the firm then this captures the firm's inability to easily substitute production across products. The covariance between these price changes and cumulative wedges reveals how joint

[4]Taking logs and differentiating gives $d\log\lambda_{21} = d\log\mu_{21} + d\log\lambda_{11}$. Since λ_{21} is constant under Cobb-Douglas demand, we have $d\log\lambda_{11} = -d\log\mu_{21}$. Then, from equation (2) and $d\log\lambda = d\log q + d\log p$, we have $d\log(q_{11}/q_{12}) = \frac{\sigma}{\sigma+1}d\log(\lambda_{11}/\lambda_{12})$.

production constraints affect misallocation — if prices rise for products with high cumulative wedges, then the scope for reallocation is limited. This leads to our main result:

Proposition 1 (Sufficient Statistics in a Simplified Economy)**.** *In this simple economy, TFP response to the markup shock to the downstream firm can be expressed as:*

$$\Delta \log TFP = \underbrace{\mathrm{Cov}_{s_1}\left(d\log p_{(1,\cdot)}, \frac{\bar{\Gamma}_1}{\Gamma_{(1,\cdot)}}\right)}_{\textit{Multi-Product Term}} - \underbrace{\left(d\log\Lambda + \tilde{\lambda}_1 d\log\mu_{21}\right)}_{\textit{Single-Product Term}},$$

where $s_1 = (\tilde{\lambda}_1, \tilde{\lambda}_2)$ *and* $\bar{\Gamma}_1 = (\tilde{\lambda}_1\Gamma_{11}^{-1} + \tilde{\lambda}_2\Gamma_{12}^{-1})^{-1}$ *is the weighted harmonic mean of cumulative wedges.*

This result provides a approach to quantifying misallocation in the presence of joint production. Rather than directly estimating technological parameters, we can infer the constraints on resource reallocation from observed price movements within firms.

2.2.4 Another Example: Response to Taste Shocks

To further illustrate how joint production affects resource allocation, we consider a case where the economy with CET joint production technology experiences taste shocks rather than markup shocks. Specifically, we examine how changes in household preferences affect aggregate TFP. Under the Cobb-Douglas utility function, when the preference weight for product 1 changes by $d\tilde{\lambda}_1$, the weight for the product 2 adjusts by $d\tilde{\lambda}_2 = -d\tilde{\lambda}_1$.

The first-order response of aggregate TFP to a taste shock can be expressed as:

$$\Delta \log TFP = -\left(1 - \frac{1}{\sigma+1}\right)\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1\right) d\log\tilde{\lambda}_1. \tag{5}$$

As before, the degree of attenuation depends critically on σ. When σ approaches infinity, firms can freely adjust their product mix to match changes in consumer preferences. When σ approaches zero, firms must maintain fixed production proportions regardless of taste shifts, eliminating any potential efficiency gains from demand-driven reallocation.

To express this in terms of observable variables, we first note that with joint production, changes in relative prices are associated with changes in production ratios from equation (2):

$$d\log(p_{11}/p_{12}) = \frac{1}{\sigma} d\log(q_{11}/q_{12}).$$

Using this equation, we can derive the relationship between relative prices and taste shocks[5]:

$$d\log(p_{11}/p_{12}) = \frac{1}{\sigma+1}\frac{1}{\tilde{\lambda}_2} d\log\tilde{\lambda}_1.$$

This allows us to decompose the TFP response:

$$\Delta\log TFP = \underbrace{-\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1\right) d\log\tilde{\lambda}_1}_{\text{Single-Product Term}} + \underbrace{\tilde{\lambda}_1\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1\right) d\log(p_{11}/p_{12})}_{\text{Multi-Product Term}}.$$

Following similar calculations as in the markup shock example, we can express the single-product term using the labor share and the multi-product term as a covariance between prices and cumulative wedges. We formalize this result in the following proposition.

Proposition 2 (Sufficient Statistics with Taste Shocks)**.** *In this simple economy, the TFP response to taste shocks can be expressed as:*

$$\Delta TFP = \underbrace{\text{Cov}_{s_1}\left(d\log p_{(1,\cdot)}, \frac{\bar{\Gamma}_1}{\Gamma_{(1,\cdot)}}\right)}_{\textit{Multi-Product Term}} - \underbrace{d\log\Lambda_L}_{\textit{Single-Product Term}},$$

where $s_1 = (\tilde{\lambda}_1, \tilde{\lambda}_2)$ *and* $\bar{\Gamma}_1 = (\tilde{\lambda}_1\Gamma_{11}^{-1} + \tilde{\lambda}_2\Gamma_{12}^{-1})^{-1}$ *is the weighted harmonic mean of cumulative wedges.*

These examples of markup shocks and taste shocks demonstrate a notable feature of our sufficient statistics approach: despite the different nature of the underlying shocks, their impact on allocative efficiency can be measured using the same statistics.

Moreover, when multiple shocks occur simultaneously, prices and factor shares reflect the combined impact of these shocks. This is particularly useful because real-world data can be considered to be generated as a consequence of compound shocks, meaning we can infer changes in allocative efficiency from observed data.

While our examples have focused on a simplified economy, many of these insights carry over to more general settings. We now turn to introducing a more general economy.

2.3 General Production Network Setup

We present our general framework to measure misallocation without imposing parametric assumptions on firms' technologies. We allow for arbitrary firm-to-firm linkages and arbitrary joint

[5]The relationship between taste shocks and relative prices can be derived as follows. Under Cobb-Douglas preferences, changes in expenditure shares directly reflect taste shocks: $d\log\lambda_{11} = \frac{d\tilde{\lambda}_1}{\tilde{\lambda}_1}$ and $d\log\lambda_{12} = -\frac{d\tilde{\lambda}_1}{\tilde{\lambda}_2}$. The relative price change is related to quantity changes through the elasticity of transformation: $d\log(p_{11}/p_{12}) = \frac{1}{\sigma}d\log(q_{11}/q_{12})$. Combining these with the relationship $d\log\lambda = d\log p + d\log q$ yields $d\log(p_{11}/p_{12}) = \frac{1}{\sigma+1}\frac{1}{\tilde{\lambda}_2}d\log\tilde{\lambda}_1$.

production technologies and, hence, heterogeneous transformation elasticities across products within firms.

Multiproduct Firms

Firm $i \in \mathcal{N}$ produces product $g \in \mathcal{G}$ and uses products $g' \in \mathcal{G}$ from other firms $j \in \mathcal{N}$ and factors (Labor, L and Capital, K) as production inputs. [6] We assume the following production set with CRS and separability between input and output functions:

$$F_i^Q\left(\underbrace{\left\{q_{ig}\right\}_{g\in G}}_{\text{outputs}}\right) = A_i F_i^X\left(\underbrace{\left\{x_{i,jg'}\right\}_{j\in N, g'\in G}}_{\text{Intermediate product } g' \text{ from } j}, L_i, K_i\right), \tag{6}$$

where A_i represents the productivity at the firm level.[7] While the model specifies Hicks-neutral productivity, this formulation can accommodate input-biased productivity.[8] However, incorporating product-specific productivity differences requires that such productivity be uncorrelated with the cumulative wedge introduced in Section 2.5 for Proposition 3 to apply.

Firms charge a product-specific markup, μ_{ig}, over its product-specific marginal cost; thus, the price is defined as $p_{ig} = mc_{ig}\mu_{ig}$.

Final Demand

Real GDP is the maximizer of a constant-returns homothetic aggregator of final uses of products: $Y = \max_{(c_{i1},\dots,c_{NG})} U\left(c_{i1},\dots,c_{NG}\right)$ subject to the budget constraint

$$\sum_{i\in\mathcal{N}}\sum_{g\in\mathcal{G}} p_{ig}c_{ig} = \sum_{f\in\{L,K\}} w_f L_f + \sum_{i\in\mathcal{N}}\sum_{g\in\mathcal{G}}\left(1 - 1/\mu_{ig}\right)p_{ig}q_{ig},$$

where w_f is the price of factor f.

Each product can be consumed by final consumers (c_{ig}) or used as an input in production by other firms ($x_{ji,g}$). The following resource constraint applies:

$$q_{ig} = c_{ig} + \sum_{j\in\mathcal{N}} x_{jig}, \quad \sum_{i\in\mathcal{N}} L_i = L, \quad \sum_{i\in\mathcal{N}} K_i = K.$$

[6]We treat factors exhibiting zero return to scale production functions; they generate production inputs without using inputs from other firms.

[7]While this formulation assumes common input intensities across different production activities within a firm, this is not a theoretical restriction but rather a measurement constraint. If data exists for separable production activities, these can be treated as if they were (joint) production activities of different firms. Since such data is not available in our application, we assume common input bundles for each firm.

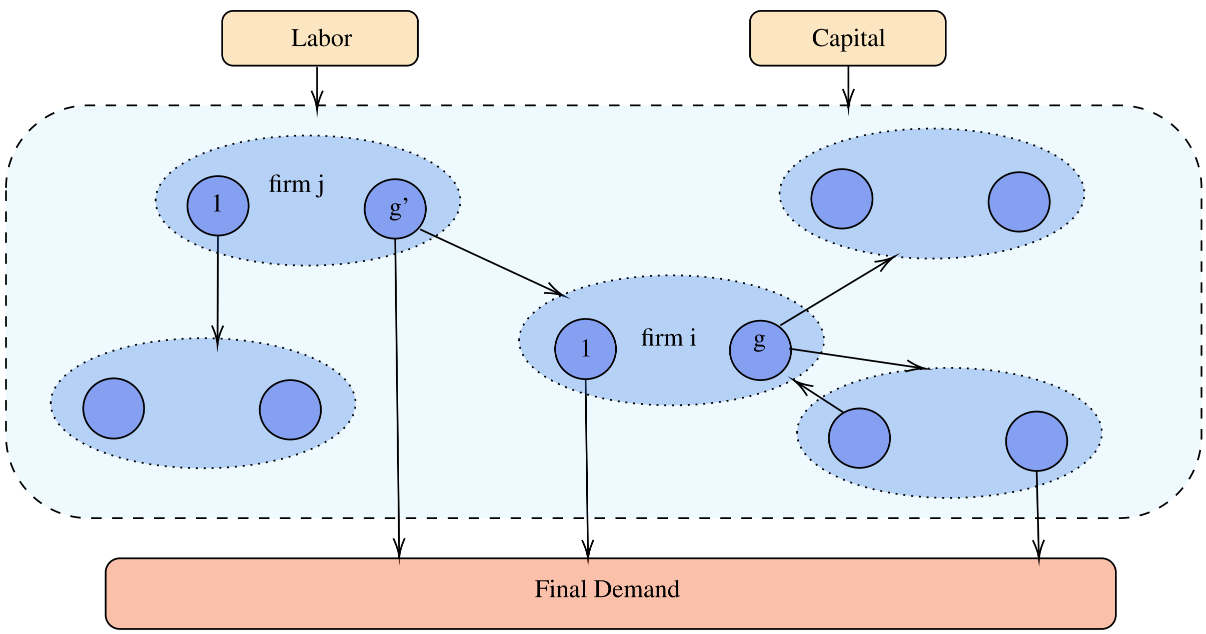
[8]Input-specific productivity can be captured by introducing a fictitious producer who purchases input j and sells to producer i using a linear technology, with Hicks-neutral shocks applied to this fictitious producer.

Figure 2 presents a stylized representation, showing the flow of products.

General Equilibrium

Given a vector of firm-level productivity, $\boldsymbol{A}$, and vector of product-level markups, $\boldsymbol{\mu}$, for all $i \in \mathcal{N}$ and $g \in \mathcal{G}$, the general equilibrium is a set of prices (p_{ig}) intermediate input choices ($x_{ijg'}$), factor input choices (L_i, K_i), output, (q_{ig}), and consumption choices (c_{ig}). As such, (i) the price of each product is equal to its markup multiplied by its marginal cost; (ii) households maximize utility under budget constraints, given prices; and (iii) markets are clear for all products and factors.

Figure 2: Graphical illustration of networks with multiproduct firms



Notes: The dashed line represents firms' universe $\mathcal{N}$, the dotted circled line represents each firm's boundary, and the circled line represents each product within a firm. The two top nodes represent factors, and the bottom node represents households. Arrows represent the direction of input flows.

2.4 Input–Output Definitions

To state our decomposition results, we introduce notation for input-output relationships at the product level.

Product-Level Input–Output Matrix

The product-level input–output matrix $\tilde{\Omega}$ is a ($\mathcal{N}\mathcal{G} + \mathcal{F}$) square matrix. Here, $\mathcal{N}$ is the number of firms, $\mathcal{G}$is the number of products, and $\mathcal{F}$ is the number of factors. $\tilde{\Omega}$ has at its ig, jg'^{th} element the expenditure share of product g' from firm j and factor $f \in \mathcal{F}$ used by firm i in production over firm i total costs (of producing all its products). The separability assumption indicates that the same expenditure share applies for all products, g, that firm i produces; thus, $\tilde{\Omega}_{ig,jg'}$ and $\tilde{\Omega}_{ig,f}$ are as follows.

$$\tilde{\Omega}_{ig,jg'} = \frac{p_{jg'} x_{i,jg'}}{\sum_{j \in \mathcal{N}, h \in \mathcal{G}} p_{jh} x_{i,jh} + \sum_{f \in \mathcal{F}} w_f x_{if}}, \qquad \tilde{\Omega}_{ig,f} = \frac{w_f x_{if}}{\sum_{j \in \mathcal{N}, h \in \mathcal{G}} p_{jh} x_{i,jh} + \sum_{f \in \mathcal{F}} w_f x_{if}}.$$

The product cost-based Leontief inverse $\tilde{\Psi}$ captures each firm-product pair's direct and indirect cost exposures through production networks. We use each $\tilde{\Psi}$ element to measure the weighted sum of all paths between two nonzero firm-product pairs.

$$\tilde{\Psi} \equiv (I - \tilde{\Omega})^{-1} = i + \tilde{\Omega} + \tilde{\Omega}^2 + \dots$$

We define the final consumption share vector, b, as follows:

$$b_{ig} = \begin{cases} \frac{p_{ig} c_{ig}}{GDP} & \text{if } i \in \mathcal{N}, g \in \mathcal{G} \\ 0 & \text{otherwise} \end{cases}$$

where $GDP = \sum_{i \in \mathcal{N}} \sum_{g \in \mathcal{G}} p_{ig} c_{ig}$. We set GDP to be the numeraire and define the product-level cost-based Domar weight, $\tilde{\lambda}_{ig}$. [9] This measures the importance of product g from firm i in final demand in two dimensions: directly when sold to final consumers, and indirectly through the production network when product g is sold to other firms and eventually reaches final consumers via downstream production networks.

$$\tilde{\boldsymbol{\lambda}}' \equiv \boldsymbol{b}' \tilde{\Psi} = \boldsymbol{b}' + \boldsymbol{b}' \tilde{\Omega} + \boldsymbol{b}' \tilde{\Omega}^2 + \dots$$

Factor shares are defined as

$$\Lambda_L = \frac{wL}{GDP}, \quad \Lambda_K = \frac{rK}{GDP}.$$

Firm-Level Aggregation

Summing over products by firms allows us to recover the firm-level cost-based Domar weight $\tilde{\lambda}_i$, which we use to compute the within-firm product-level Domar weight share s_{ig}:

$$\tilde{\lambda}_i = \sum_{g \in \mathcal{G}} \tilde{\lambda}_{ig}, \quad s_{ig} = \frac{\tilde{\lambda}_{ig}}{\tilde{\lambda}_i}.$$

[9] We denote $\tilde{\Lambda}_f$ with $f \in \{L, K\}$.

Finally, we define firm-level aggregate markup as follows:

$$\mu_i = \frac{\text{sales of } i}{\text{total cost of i}},$$

2.5 Cumulative Wedges

Building on our insights from the simplified economy, we now generalize the concept of cumulative wedges to arbitrary production networks with multiproduct firms. The example shows that products can face different cumulative distortions depending on their downstream supply chain. We now formalize this notion for arbitrary production networks.

Definition 2 (Cumulative Wedge)**.** For product g of firm i, the cumulative wedge is defined as:

$$\Gamma_{ig} \equiv \underbrace{\tilde{\lambda}_{ig}/\lambda_{ig}}_{\text{downstream wedges}} \times \underbrace{\mu_{ig}}_{\text{own wedge}},$$

where λ_{ig} denotes sales share of firm i's product g over GDP.

The cumulative wedge summarizes the cumulative distortion in the downstream supply chain of product g sold by firm i. In efficient economies with no markups, the product cost-based Domar weight equals observed sales shares, generating a cumulative wedge equal to one for all products and firms. Conversely, in an inefficient economy, a portion of the indirect demand transmitted from downstream firm-product pairs to upstream firm-product firms is absorbed as profit by downstream firms. This effect accumulates in each supply chain transaction upstream until indirect demand reaches product g sold by firm i; thus, the sales share of a product is smaller relative to an efficient economic outcome. Therefore, the larger the ratio, the greater the cumulative wedges in the downstream supply chain.

For our aggregation result, we compare distortions across products within the same firm. We do this by defining the weighted harmonic mean of cumulative wedges for firm i:

$$\bar{\Gamma}_i = \mathbb{E}_{s_i}[\Gamma^{-1}_{(i,g)}]^{-1}$$

where s_i represents the vector of within-firm cost-based Domar weight shares.

Cumulative Wedges in a Simplified Example

To illustrate how cumulative wedges capture distortions, we revisit our simplified economy composed of two firms and a representative household. We normalize GDP to 1; hence, in this setup, sales (shares) to final consumption are $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ for products 1 and 2 respectively; however, firm 1's sales of product 1 are reduced by the markup charged by firm 2, which is $\tilde{\lambda}_1/\mu$.

The product cost-based Domar weights are $\tilde{\lambda}_1$ for both products 1 and 2. In matrix notation, the value-added share vector (b) and the product cost-based input-output matrix ($\tilde{\Omega}$) are:

$$b = \begin{bmatrix} \tilde{\lambda}_1 \\ 0 \\ \tilde{\lambda}_2 \\ 0 \end{bmatrix}, \quad \tilde{\Omega} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where the matrix and vector components are arranged in the following order: product 1 and 2 of firm 1, firm 2, and labor. Therefore, the product cost-based Domar weights can be computed as:

$$\begin{aligned} \tilde{\lambda}' &= b' + b'\tilde{\Omega} + b'\tilde{\Omega}^2 + \dots, \\ &= \underbrace{[\tilde{\lambda}_1, 0, \tilde{\lambda}_2, 0]}_{\text{Final demand}} + \underbrace{[0, \tilde{\lambda}_1, 0, 0]}_{\text{Indirect demand}} . \\ &= [\tilde{\lambda}_1, \tilde{\lambda}_1, \tilde{\lambda}_2, 0]. \end{aligned}$$

These weights represent the counterfactual sales shares if markups were removed while keeping expenditure shares constant. Following the definition, the cumulative wedge for firm 1's products is:

$$\Gamma_{11} = \frac{\tilde{\lambda}_1}{(\tilde{\lambda}_1/\mu)}\mu = \mu^2, \qquad \Gamma_{12} = \frac{\tilde{\lambda}_1}{\tilde{\lambda}_1}\mu = \mu.$$

The markup of product 2 from firm 1 and the product from firm 2 equal μ. Comparatively, product 1 from firm 1 has a larger cumulative wedge of μ^2 than that of product 2, reflecting both the product's own markup and the downstream distortions the product faces. In this case, product 1 from firm 1 generates a distortion by charging a markup and is subject to an additional distortion through downstream production networks because firm 2 uses the marked-up input in its production.

The next section shows how these wedges enter our main aggregation result for arbitrary production networks with multiproduct firms.

2.6 Aggregation Theorem with Multiproduct Firms within Production Networks

To allow for joint production and derive sufficient statistics, this section generalizes the concept of an allocation matrix introduced by Baqaee and Farhi (2020).

Let $\mathcal{X}$ be an $(\mathcal{N} + \mathcal{F}) \times (\mathcal{N}\mathcal{G} + \mathcal{F})$ admissible input allocation matrix; the rows are buyer firms, and the columns are seller-product pairs. Each element, $\mathcal{X}_{ijg} = \frac{x_{ijg}}{q_{jg}}$, is the share of the output of product g produced by firm j that firm i uses as a production input.

A productivity shock ($\mathrm{d}\log A$) and a markup shock ($\mathrm{d}\log \mu$) effect in real GDP, Y, can be decomposed into changes in the distribution of $\mathcal{X}$ ($\mathrm{d}\mathcal{X}$), holding productivity constant, and a pure change in productivity ($\mathrm{d}\log A$) for a given fixed allocation matrix $\mathcal{X}$. In vector notation:

$$\mathrm{d}\log\mathcal{Y} = \underbrace{\frac{\partial\log\mathcal{Y}}{\partial\mathcal{X}}\mathrm{d}\log\mathcal{X}}_{\Delta\text{ Allocative Efficiency}} + \underbrace{\frac{\partial\log\mathcal{Y}}{\partial\log A}\mathrm{d}\log A}_{\Delta\text{ Technology}}. \tag{7}$$

We now present a decomposition of changes in aggregate TFP that considers multiproduct firms and arbitrary production networks with product-level distortions.

Proposition 3 (Growth Accounting in Networks with Multiproduct Firms)**.** *To the first order, aggregate TFP can be decomposed into technology and allocative efficiency terms as follows:*

$$d\log TFP = \underbrace{\underbrace{\sum_i \tilde{\lambda}_i \mathrm{Cov}_{s_i}\left(d\log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right)}_{\textit{Multi-Product Term}} \underbrace{- \sum_f \tilde{\Lambda}_f d\log\Lambda_f - \sum_i \tilde{\lambda}_i d\log\mu_i}_{\textit{Single-Product Term}}}_{\Delta\textit{ Allocative Efficiency}} + \underbrace{\sum_i \tilde{\lambda}_i d\log A_i}_{\Delta\textit{Technology}},$$

where $d\log p_{(i,\cdot)} = (d\log p_{i1}, ..., d\log p_{iG})$ denotes the vector of price changes, $\Gamma_{(i,\cdot)} = (\Gamma_{i1}, ..., \Gamma_{iG})$ represents the vector of cumulative wedges for firm i's products, and $\bar{\Gamma}_i = \mathbb{E}_{s_i}[\Gamma^{-1}_{(i,g)}]^{-1}$ is the weighted harmonic mean of cumulative wedges.

Appendix F presents the proof. The change in aggregate TFP can be decomposed into technology and allocative efficiency terms. The technology term represents a weighted average of changes in firm-level Hicks-neutral productivity using cost-based Domar weights. The allocative efficiency term is further decomposed into a multiproduct firm term, a change in aggregate factor shares, and firm-level average markup changes.

The multiproduct term captures the allocative efficiency implications of firm-level product mix adjustments. When a firm adjusts its product mix, the relative prices of its products change to reflect the reallocation costs imposed by technological constraints. These price changes interact with existing distortions: if prices rise more for products with higher cumulative wedges (captured by a positive covariance), technological constraints limit reallocation precisely where it would be most beneficial for efficiency. In this general setting, these opportunity costs vary across firms and product pairs, reflecting differences in the curvature of their production possibility frontiers. For example, an oil refinery producing gasoline and diesel may face different constraints when adjusting its production mix than a dairy farmer producing milk and meat.

To calculate the aggregate effect across the economy, we sum these firm-level covariances using Domar weights, which indicate its macroeconomic importance. This aggregation allows us to quantify the overall impact of product mix changes on allocative efficiency in the economy.

Regarding the single product term, which consists of factor shares and firm-level markup, if the initial equilibrium is inefficient, the products charging markups are underproduced relative to an efficient economy. Improving the allocation involves reallocating resources to a more distorted part of the economy, such as firms' product pairs that charge relatively high markups. A decrease in factor shares implies reallocating resources to the portion of the economy that has relatively high markups; however, if the change in factor share is due to a change in markup, this is a mechanical change and does not imply reallocation. Therefore, the contribution of the change must be purged, which the firm-level markup term captures. The factor shares and firm-level markup terms are proposed by Baqaee and Farhi (2020). Both terms are valid under a joint production approach and, together with the multiproduct term this work introduces, constitute allocative efficiency.

Relation to Existing Aggregation Theorems

Proposition 3 nests existing aggregation theorems for production networks as a special case.

Corollary 1 (Baqaee and Farhi (2020))**.** *If no firms engage in joint production and impose the same markup on all their products (the single-product firm assumption), then to a first order, aggregate TFP growth can be decomposed into technology and allocative efficiency terms as follows.*

$$d \log TFP = \underbrace{\sum_{i \in \mathcal{N}} \tilde{\lambda}_i d \log A_i}_{\text{Technology}} \underbrace{- \sum_{i \in \mathcal{N}} \tilde{\lambda}_i d \log \mu_i - \sum_{f \in \mathcal{F}} \tilde{\Lambda}_f d \log \Lambda_f}_{\text{Allocative Efficiency}} .$$

The proof follows from the fact that the covariance term from Proposition 3 is zero because the changes in marginal cost and markup for all products within a firm are equal.

Our approach quantifies misallocation through the multiproduct channel by measuring deviations from the single-product, single-markup assumption when product-level data are available; if this assumption holds, the multiproduct term becomes zero. The assumption of uniform marginal costs and markups is unlikely to hold in practice; however, its quantitative relevance remains an empirical question. Our decomposition quantifies the extent to which this assumption is violated and isolates the impact of existing misallocation literature.

Finally, without markups, when prices equal marginal costs, allocative efficiency converges to zero. In this case, all aggregate TFP changes are attributed to technology, aligning with Hulten (1978).

Corollary 2 (Hulten (1978)). *Growth Accounting in an Efficient Economy:*

$$d \log TFP = \underbrace{\sum_{i \in \mathcal{N}} \tilde{\lambda}_i d \log A_i}_{\text{Technology}} .$$

The proof follows from the fact that the markup is always 1, the markup change term is 0, and the sum of factor shares is always 1. Therefore, the sum of factor changes is always 0, and the covariance of the multiproduct term is 0 because $\Gamma_{(i,\cdot)} = (\Gamma_{i1}, ..., \Gamma_{iG})$ *are all 1 in an efficient economy.*

Proposition 3's formula converges to Hulten's theorem when the economy is efficient. Measured aggregate TFP growth equals the Domar weighted sum of firm-level productivity changes.

3 Empirical Evidence on Joint Production

In Section 2, we developed a theoretical framework that accommodates multiproduct firms using joint production technologies. When production technologies across products within firms are separable, our framework collapses to existing aggregation theorems that treat each product as a separate firm. The literature on multiproduct firms often assumes such product line independence (Bernard et al. (2010, 2011); Hottman et al. (2016); Mayer et al. (2021)).

We exploit geographic variation in firms' exposure to local demand shocks to test for joint production. Using Chilean firm-to-firm transaction data, we leverage that each product within a firm has its own set of buyers in different locations. When some locations experience negative demand shocks due to COVID-19 lockdowns, this creates product-specific variation in demand within firms. Following Hall (1973); Ding (2023), we first examine whether negative shocks to the demand for one product affect the production and pricing of other products within the same firm. We then exploit the same quasi-experimental variation to estimate the elasticity of transformation between products, providing a quantitative measure of these technological linkages.

3.1 Partial Equilibrium Setting for Reduced-Form Regression

Consider a firm operating with the joint production technology characterized by the CET function introduced in Section 2 where its cost function can be expressed as:

$$C(q_1, \ldots, q_N) = \frac{P_M}{A} \left(\sum_{i=1}^{N} \left(\frac{q_i}{a_i} \right)^{\frac{\sigma+1}{\sigma}} \right)^{\frac{\sigma}{\sigma+1}} \quad \sigma > 0, \tag{8}$$

where q_g represents the output of product g, and P_M denotes a composite input price index combining inputs, potentially any combination of intermediate inputs, labor, and capital. The parameter σ represents the elasticity of transformation between outputs in production. For each product g, we assume a standard isoelastic demand with cross-price elasticity equal to zero[10]:

$$q_g = D_g \, p_g^{-\theta_g}, \quad \theta_g > 1, \tag{9}$$

where $D_g > 0$ is the demand shifter for product g, θ_g is the own-price elasticity, and p_g is the price of product g. We allow a reduced-form wedge μ_g such that $p_g = \mu_g \times \frac{\partial C}{\partial q_g}$. We assume changes in D_g are uncorrelated with changes in μ_g; therefore, exogenous variation in D_g shifts the firm's output choice for good g without directly affecting the wedge.

Proposition 4 (Within-Firm Demand Shock Spillover via Joint Production)**.** *Consider a negative demand shock to product k ($d \log D_k < 0$). For products $g \neq k$, this leads to:*

(i) Quantity response: $d \log q_g < 0$,

(ii) Price response: $d \log p_g > 0$.

When product k experiences a negative demand shock, its production is dampened, and hence, the opportunity cost of producing other products increases through the concave production possibilities frontier. These changes in opportunity costs translate into higher marginal costs for other products, leading firms to raise their prices and reduce their quantities.

In the extreme, if $\sigma = \infty$ in (8), product lines are independent. Then, a shock to D_k *does not* alter mc_g for $g \neq k$, meaning that there is *no cross-product spillover*. Formally:

Lemma 1 (Independent Product lines)**.** *When products are perfectly separable in production ($\sigma = \infty$), a demand shock to product k has no effect on other products $g \neq k$:*

$$d \log q_g = 0, \quad d \log p_g = 0, \quad \forall \, g \neq k.$$

This lemma corresponds to the standard assumption in the multiproduct firm literature, where product lines operate independently.

Complete proofs of Proposition 4 and Lemma 1 are provided in Appendix D. Appendix B presents examples of market structures that generate our assumed relationships between wedges and demand shifters.

3.2 Data and Empirical Strategy

We use data from the Chilean Internal Revenue Service (SII), covering all formal firms in Chile.[11] We then employ monthly data from January 2019 to December 2021 to test for joint production.

[10] Hottman et al. (2016) studies settings where firms' products compete for the same customers or in the same markets, generating non-zero cross-price elasticities through cannibalization. Our empirical strategy focuses on firms where the buyers of the shocked product do not overlap with the buyers of other products whose responses we examine.

[11] This study was developed within the scope of the research agenda conducted by the Central Bank of Chile (CBC) in economic and financial affairs of its competence. The CBC has access to anonymized information from various public and private entities by virtue of collaboration agreements signed with these institutions. To secure the privacy of workers and firms, the CBC mandates that the development, extraction, and publication of the results should not allow

The SII provides detailed information on firm-to-firm transactions through electronic tax documents. This dataset, which captures every product, quantity, and price traded between formal Chilean firms, contains data on over 15 million unique firm-specific product descriptions.[12] We divide the value traded over its quantity to obtain average prices, which will be monthly on the empirical evidence section and yearly on our quantification exercise.[13]

To test the model's predictions about spillovers across products, we exploit geographic variation in COVID-19 lockdowns across Chilean counties in March 2020 as exogenous demand shocks. These lockdowns represent negative shocks to the demand shifters D_g in our theoretical framework, allowing us to examine how firms adjust their production of other products in response.

COVID-19 Lockdowns in Chile

The Chilean government implemented county-specific lockdowns beginning in March 2020. We focus on this initial period to ensure the shock was unexpected. Figure 3 illustrates the spatial heterogeneity of these lockdowns.

---

the identification, directly or indirectly, of natural or legal persons. Officials of the CBC processed the disaggregated data. All the analysis was implemented by the authors and did not involve nor compromise the Chilean IRS. The information contained in the databases of the Chilean IRS is of a tax nature originating in self-declarations of taxpayers presented to the Service; therefore, the veracity of the data is not the responsibility of the Service.

[12]The specific invoice variable is called "detail", which is inherently firm-specific and can differ between firms even for the same product. For example, one supermarket might declare selling "Sprite can 330cc" while another declares selling "Sprite 330". This variation across sellers does not affect our analysis in this section as we do not compare identical products across firms.

[13]In Appendix A.1, the distribution of the number of products is provided

Figure 3: Distribution of early Covid-19 lockdown in Chile

Notes: Lockdown counties as of March 2020 are red; all others are gray.

We exploit this geographic variation as a source of demand shocks to intermediate input transactions, treating lockdowns as negative demand shocks (reductions in D_g) from buyers in lockdown areas to their suppliers in non-lockdown areas. Appendix A.1 validates this approach, demonstrating that firms in lockdown areas reduced their intermediate input purchases by approximately 20%.

Empirical Evidence for Joint Production

To investigate how demand shocks to one product affect the production of other products within firms, we focus on shocks to firms' main products (defined by highest sales from January 2019 to December 2021). We classify a firm as experiencing a main product demand shock if at least one buyer of its main product is located in an area that implemented a March 2020 lockdown.

To quantify this effect, we study the impact of demand shocks to a firm's main product on the production of its other products using an event-study specification for all products $g \neq m$:

$$\log X_{igt} = \sum_{j=-11}^{10} \beta_j D_{i,t-j} + FE_{ig} + FE_t + \varepsilon_{igt}, \tag{10}$$

where X_{igt} represents either the quantity or price of product g for firm i at time t. $D_{i,t-j}$ is a treatment indicator equal to one if firm i was treated j months ago. FE_{ig} and FE_t are firm-product and time fixed effects, respectively. The coefficients of interest are β_j, which capture the effect of the main product's demand shock on other products' quantities or prices at different time points relative to the shock.

To obtain unbiased estimates of β_j, the treatment indicator $D_{i,t-j}$ must be conditionally orthogonal to the error term ε_{igt}. A concern is that supply-side shocks could be correlated with the lockdown if suppliers and main product buyers are located in the same area, potentially confounding our results. To address this issue and isolate the impact of demand shocks from the main product while ruling out direct supply shocks, we impose the following restrictions:

1. **Firm Location**: The firm itself is not located in an area under lockdown.

2. **Supplier Location**: The firm's direct suppliers are not subject to lockdown shocks.

3. **Buyer Location for Product** g: None of the buyers of product g are located in lockdown areas.[14]

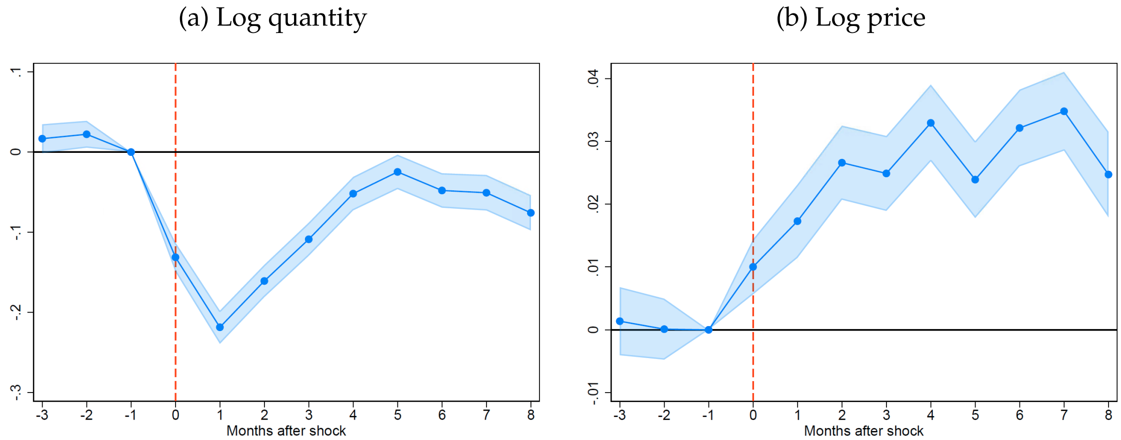
Restrictions 1 and 2 help eliminate direct supply-side effects, ensuring that any observed changes in production are not due to supply disruptions that affect the firm or its suppliers. Restriction 3 ensures that product g is not subject to a direct demand shock, allowing us to attribute any changes in its production to the demand shock that affects the main product m. It also ensures that buyers of the main product and product g are different, eliminating the impact of shocks to the main product on product g through demand complementarities and justifying equation (9).

Our treatment group comprises firms meeting these conditions with main products experiencing March 2020 demand shocks. The control group includes firms satisfying the conditions but whose main products remained unaffected by lockdowns. We drop small seller firms, those that trade below the median value transacted, by doing so we still keep above 99% of value transacted but drop potentially small share providers. Figure 4 presents the regression results.[15].

[14] Within the same firm, each product typically has its own set of buyers. As a result, when buyers of one product are affected by the lockdown, buyers of other products may remain unaffected. This distinction is further detailed in Figure A1 of Appendix A.1.

[15] A comparison of observable characteristics between the treatment and control groups is provided in Table A2 of Appendix A.1. Appendix Table A3 provides sensitivity analyses using the Difference-in-Differences (DiD) specification.

Figure 4: The effects of demand shocks to the main product on the production of other products within the firm



Notes: Standard errors are clustered at the firm-county level, and the error bands represent 95% confidence intervals. The X-axis represents the time to treat, with 0 denoting March 2020, when the main product experienced the demand shock. The other values indicate the number of months before or after this event.

The results support our joint production setting with finite elasticity of transformation. First, the pre-shock stability of quantities and prices between treatment and control groups supports the parallel trends assumption and the unanticipated nature of initial closures. Second, consistent with Proposition 4(i), we find a significant above 10 % decrease in non-main product quantities following main product demand shocks, with persistent effects. Third, aligning with Proposition 4(ii), we observe sustained price increases for other products. These combined quantity and price spillovers match our theoretical predictions under finite σ and reject the product line separability hypothesis.

Discussion on Other Within-Firm Spillover Mechanisms

Our findings of negative quantity spillovers and positive price spillovers across products within firms contrast with several alternative mechanisms in the literature:

First, Almunia et al. (2021) propose a model of diminishing returns to scale or firm-specific factors at the firm level to explain how a decline in domestic demand in Spain affects exports. Their model predicts that when there is a negative demand shock for a product in one market, firm-specific factors are reallocated to another product in another market, positively affecting the production of the same product in other markets. This prediction contrasts with our findings, which show negative spillovers across products within the same firm.

Second, Ding (2023) focuses on industries that share knowledge-intensive inputs to examine joint production effects in the US using Census data. This paper, like ours, predicts that when a product faces a negative demand shock, it negatively affects other products. The study interprets the model prediction as knowledge spillovers across industries sharing intangible inputs; however, knowledge spillovers are unlikely to explain our results. The differences in time horizon (five years vs. monthly data) and research and development (R&D) intensity (Chile's R&D spending is less than one-tenth that of the US as a percentage of GDP) limit its applicability to our context.

3.3 Estimating the Elasticity of Transformation

Finally, using the same sample of firms and product pairs from our event study analysis, we estimate the elasticity of transformation σ. Beyond providing direct evidence of joint production, this parameter is used for markup estimation to implement Proposition 3 and a counterfactual analysis in Section 6. Throughout this section, Δ refers to the yearly difference.

For each firm i, we focus on pairs of its main product (denoted m) that experienced a lockdown shock and its other products (denoted g) whose buyers were not in lockdown areas. Let p_{im} and p_{ig} be their prices, and q_{im} and q_{ig} be their respective quantities. We regress the change in the relative price on the change in the relative quantity:

$$\Delta \log\left(\frac{p_{ig}}{p_{im}}\right) = \alpha + \beta \Delta \log\left(\frac{q_{ig}}{q_{im}}\right) + \eta_{gt} + \xi_{igt}. \tag{11}$$

where β is the main coefficient of interest, and η_{gt} denotes product-year fixed effects that control for supply and unobserved demand shocks common to products within each product category.

The error term $\xi_{i,g,m,t}$ may contain supply-side factors such as wedge changes μ_g or technology shifters changes a_g that could correlate with relative quantities.[16]

To address this endogeneity concern, we employ the demand shock from our event study as an instrument: the exposure of the main product's buyers to local lockdowns. This demand shock remains orthogonal to supply-side productivity or markup shifts that might affect relative quantities through the production technology. Under our constant elasticity demand specification in equation (9), demand shifter D_m changes are uncorrelated with wedge and technology shocks, satisfying the exclusion restriction. [17]

[16]For any two products g and h within the same firm, the ratio of marginal costs satisfies, $\frac{mc_g}{mc_h} = \left(\frac{q_g/a_g}{q_h/a_h}\right)^{1/\sigma}$. Taking logarithms with wedge adjustment and differences across time, yields:

$$\Delta \log\left(\frac{p_g}{p_h}\right) = \Delta \log\left(\frac{\mu_g}{\mu_h}\right) + \frac{1}{\sigma} \Delta \log\left(\frac{q_g}{q_h}\right) - \frac{1}{\sigma} \Delta \log\left(\frac{a_g}{a_h}\right).$$

[17]A concern may arise regarding the potential correlation between relative quantities and relative markups within a firm. Marshall's Second Law of Demand, a standard property of variable-markup demand systems including the Kimball aggregator (Kimball, 1995; see Matsuyama, 2025 for a recent review), implies that higher relative demand could lower the demand elasticity, so that the markup ratio μ_g/μ_m may co-move with the relative quantity q_g/q_m. Since our instrument $Z = \Delta \log D_m$ shifts relative quantities through within-firm reallocation, any such co-movement would enter the IV moment. Appendix C formalizes this argument and shows that the resulting bias, if present, is one-sided: as long

Estimation Results

Table 1 reports estimates of β under different specifications, each using one-year differences and the same sample restrictions from the event study.

Table 1: Estimating σ

	(1)	(2)	(3)
β	0.865***	1.207***	0.630***
	(0.063)	(0.259)	(0.137)
Implied σ	1.156	0.828	1.589
Time FE	N	Y	Y
Industry FE	N	N	Y
F-stat first stage	699.22	52.00	98.56

Notes: The Table reports the results of estimating equation (11), clustered at the firm-municipality level. Column (1) reports the 2SLS estimation without fixed effects, while columns (2) and (3) report results by 2SLS with time and industry fixed effects. Three stars indicate statistical significance at the 1% level.

Across columns, the implied σ values range between about 0.8 and 1.6. These estimates reject the hypothesis of full separability ($\sigma \to \infty$), indicating concave curvature in firms' production possibilities frontiers.

4 Construction of Sufficient Statistics

Having established the presence of joint production, we implement our sufficient statistics framework developed in Section 2 using a dataset from the Chilean Internal Revenue Service (Servicio de Impuestos Internos, SII). As discussed in Section 3, the dataset primarily relies on electronic tax invoices, which provide detailed records of all firm-to-firm transactions, including product descriptions, quantities, and prices. These invoices allow us to observe the complete structure of firm-to-firm relationships and compute firm-specific product shares.

To construct a full input output matrix, and cumulative wedge, we additionally use tax accounting declarations, which provide monthly data on each firm's revenue and input expenditures, including capital and labor costs. A key advantage of the SII data is its use of unique identifiers for firms and workers, which allows individual and firm data to be merged across datasets. We utilize four distinct sources from SII.

The first is the value-added tax form, which includes gross monthly firm sales, materials expenditures, and investment.

as MSL holds, $\hat{\beta}$ would be biased downward and $\hat{\sigma}$ correspondingly upward. Given that the literature conventionally assumes $\sigma = \infty$, our estimates therefore provide an upper bound for σ, serving as a conservative estimate even in the presence of variable markups.

Second, the SII provides information from a matched employer–employee census of Chilean firms from 2005 to 2022. Specifically, firms must report all payments to individual workers, including the sum of taxable wages, overtime, bonuses, and any other labor earnings for each fiscal year. All legal firms must report to the SII; thus, the data cover the total labor force with a formal wage contract, representing roughly 65% of employment in Chile. For any given month, it is possible to identify an individual worker's employment status, their average monthly labor income that year, a monthly measure of total employment, and the distribution of average monthly earnings within the firm.

Third, income tax form data includes yearly information on all sources of a firm's income and expenses. This form allows for computing every individual's actual tax payments for each year. Details on sales and employment are available on this form; however, we use only data on capital stock for each firm and year. This approach allows us to build perpetual inventories using data from the monthly F22 form. We obtain the user cost of capital by multiplying nominal capital stock by the real rental rate of capital, which is built using publicly available data. We use the 10-year government bond interest rate minus expected inflation plus the external financing premium. Finally, we use the capital depreciation rate from the LA-Klems database.

Fourth are electronic tax documents from 2016 onward. These documents provide information on each product (price and quantity) traded domestically or internationally with at least one Chilean firm. We only use domestic transactions and observe the firm-to-firm transactions and a firm's sales (including firm-to-firm and firm-to-consumer sales). We compute firm-specific product shares for firm-to-firm transactions and assume that their distributions are equivalent to firm-to-consumer transactions to recover the complete distribution of firm sales by product. Each firm-to-firm transaction includes a "detail" column that records the name of each product.

Building on the data cleaning process described in Section 3, we process the data to construct product code-level output and input-price indices for each firm using standard Tornqvist indices. We aggregate products into a 581 product-code identifier (CCIF) to facilitate comparison between firms, allowing us to estimate product production functions that use the same product across firms.

4.1 Data Cleaning and Implementation Strategy

We begin the data processing by applying filters to the raw data to obtain the final database for empirical analysis. We define a firm as a taxpayer with a tax ID, positive sales, positive materials, positive wage bill, and capital for any given year. We exclude firms that hire less than two employees a year or have capital valued below US$20 in a year. All variables are winsorized at the 1% and 99% levels to mitigate measurement error.

We selected 2016 as the base year for price indices because it was the first year we observed prices for firm-to-firm transactions. This method is widely recognized for estimating aggregate production functions at the firm or plant level when price data is accessible. We use crosswalks developed at the Central Bank of Chile (Acevedo et al. (2023)) to address the challenge of product aggregation (from around 15 million products to 581 product codes). We create aggregated product-level quantity produced and material usage indices, matching product descriptions and characteristics to ensure consistency across firms and over time.

4.2 Construction of Sufficient Statistics

We measure five distinct objects to implement the growth accounting framework that includes the multiproduct channel: (1) product-level cost-based Domar weights $\tilde{\boldsymbol{\lambda}}$, (2) product-firm level price indices, (3) product-level markups μ, (4) cumulative wedges, and (5) aggregate objects. We discuss each of these in the following subsection.

4.2.1 Product-Level Cost-Based Domar Weights

The product cost-based Domar weights can be calculated using the following equation:

$$\tilde{\boldsymbol{\lambda}}' \equiv \boldsymbol{b}'\tilde{\Psi} = \boldsymbol{b}' + \boldsymbol{b}'\tilde{\Omega} + \boldsymbol{b}'\tilde{\Omega}^2 + \dots$$

To compute these weights, we must measure value-added shares ($\boldsymbol{b}$) and the input–output matrix ($\tilde{\Omega}$). We measure these objects directly from the data.

Final expenditure shares ($\boldsymbol{b}$) are represented by a vector of dimension $(\mathcal{N}\mathcal{G} + \mathcal{F}) \times 1$. Here, $\mathcal{N}$ is the number of firms, $\mathcal{G}$ is the number of products, and $\mathcal{F}$ is the number of factors. The first $\mathcal{N}\mathcal{G}$ entries are calculated as the residual between a firm product's total sales and its intermediate sales to other firms (measured from the firm-to-firm data). This approach provides a theory-consistent measure of final expenditures. The final $\mathcal{F}$ entries are set to zero because households do not directly purchase factors. Using firm-to-firm records and factor expenditures, we construct the input–output matrix $\tilde{\Omega}$ at the product-firm level.

Specifically, we compute an annual cost-based input–output matrix by product. We calculate the denominator of each element (indexed by ig, jg') by summing a firm's purchases from all its suppliers, its wage bill, and its capital multiplied by the relevant user cost rental rate of capital. The last two elements of the matrix have wage bills and capital expenditures as their numerators.

The resulting $\tilde{\Omega}$ is a $(\mathcal{N}\mathcal{G} + 2) \times (\mathcal{N}\mathcal{G} + 2)$ matrix that can be expressed as follows:

$$\tilde{\Omega} = \left[\begin{array}{ccc|cc} \tilde{\Omega}_{11,11} & \cdots & \tilde{\Omega}_{11,\mathcal{NG}} & \tilde{\Omega}_{11,\mathcal{NG}+1} & \tilde{\Omega}_{11,\mathcal{NG}+2} \\ & \ddots & & & \\ \tilde{\Omega}_{\mathcal{NG},11} & & \tilde{\Omega}_{\mathcal{NG},\mathcal{NG}} & \tilde{\Omega}_{\mathcal{NG},\mathcal{NG}+1} & \tilde{\Omega}_{\mathcal{NG},\mathcal{NG}+2} \\ \hline 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{array}\right].$$

Based on the separability assumption, the same expenditure share applies to all products g that firm i produces. The expressions for $\tilde{\Omega}_{ig,jg'}$ and $\tilde{\Omega}_{ig,f}$ are as follows:

$$\tilde{\Omega}_{ig,jg'} = \frac{p_{jg'} x_{i,jg'}}{\sum_{j \in \mathcal{N}, h \in \mathcal{G}} p_{jh} x_{i,jh} + \sum_{f \in \mathcal{F}} w_f x_{if}}, \qquad \tilde{\Omega}_{ig,f} = \frac{w_f x_{if}}{\sum_{j \in \mathcal{N}, h \in \mathcal{G}} p_{jh} x_{i,jh} + \sum_{f \in \mathcal{F}} w_f x_{if}}.$$

Factors do not require inputs; thus, the last row of the matrix is zero.

After calculating the product-level cost-based Domar weights, we sum them for the same firms to compute the firm-level cost-based Domar weights and their shares. These will be inputs for Proposition 3.

$$\tilde{\lambda}_i = \sum_{g \in \mathcal{G}} \tilde{\lambda}_{ig}, \quad s_{ig} = \frac{\tilde{\lambda}_{ig}}{\tilde{\lambda}_i}.$$

4.2.2 Product-Firm Level Price Indices

We observe prices for each transaction and aggregate them into 581 product categories. We construct two types of price indices: output and input price indices. We compute firm-product-specific annual price indices for the output price index, which is an input to sufficient statistics that deflates product output for production function estimation. The original data are at the "detail" product level, which we aggregate to a Tornqvist index for each of the 581 product categories the firm owns. Specifically, we construct the following price index:

$$\Delta \log P_{igt} = \sum_{d \in g} \frac{s_{idt} + s_{idt-1}}{2} \Delta \log P_{idt},$$

where d is the detailed category belonging to the upper product category. $\Delta \log P_{idt}$ is the price change, and s_{idt} is the share at time t in the continuing products in category g. We construct our price index with 2016, the starting year of the data, as the base year. We also construct an input price index to deflate material costs for production function estimation. We define one aggregate index per firm because aggregate materials are used as inputs in production function estimation. The construction method is the same as for the output price index.

4.2.3 Cumulative Wedges

To construct the cumulative wedge measure, we need product cost-based Domar weights, product sales shares, and product markups :

$$\Gamma_{ig} \equiv \underbrace{\frac{\tilde{\lambda}_{ig}}{salesshare_{ig}}}_{\text{Downstream wedge}} \times \underbrace{\mu_{ig}}_{\text{own markup}} .$$

As discussed in Section 2, the ratio of the cost-based Domar weight to the sales share recovers the cumulative wedge accumulated downstream of a product, leaving the own markup μ_{ig} to be estimated. Recovering own markups requires taking a stand on the firm's production technology, and the literature offers several approaches that rest on different identifying assumptions. As our baseline specification, we estimate product-level markups following the methodology of Dhyne et al. (2022), which we adapt to a joint-production environment using a CET parametric production function (details provided in Appendix D); the implementation relies on our transformation-elasticity estimates from Section 3. Chilean invoice data, which record both prices and quantities, let us construct quantity-based output elasticities and thereby circumvent the identification challenges in markup estimation highlighted by Bond et al. (2021a).

As an alternative, one may instead adopt an accounting approach, in which markups are assumed homogeneous within firms and computed as the ratio of firm-level sales to firm-level costs (as in single-product firm models such as Baqaee and Farhi (2020)), so that within-firm variation in cumulative wedges arises solely from downstream wedges. We report this accounting-markup version as a robustness check in Appendix E and find that it yields qualitatively similar aggregate results.

4.2.4 Aggregate Objects

In addition to product cost-based Domar weights and cumulative wedges, we must measure aggregate objects to implement the sufficient statistics presented in Proposition 3. In particular, Y, L, K, Λ_L, and Λ_K denote aggregate value-added, employment, capital, and labor and capital shares, respectively. We measure Y, L, and K as the sum of value added, employment, and capital, respectively, for all firms in the economy. Factor shares of GDP, Λ_L and Λ_K, are measured as total compensation and capital with user cost of capital divided by GDP. Real GDP is calculated by deflating GDP with the official GDP deflator.

5 Application: Decomposing Aggregate TFP Growth

This section applies Proposition 3 to decompose aggregate TFP growth in Chile. Our analysis covers 2016 to 2022, a period of stagnant productivity growth. This productivity trend aligns with the pattern of productivity stagnation observed in Chile using different computation methods.[18] We begin with the standard single-product benchmark and then show how the decomposition changes once we account for joint production within multiproduct firms.

We begin by presenting results using the standard assumption in the literature of single-product firms. If firms produce a single product, then Corollary 1 applies:

$$d\log TFP = \underbrace{-\sum_{i\in\mathcal{N}}\tilde{\lambda}_i d\log\mu_i - \sum_{f\in\mathcal{F}}\tilde{\Lambda}_f d\log\Lambda_f}_{\Delta\text{ Allocative Efficiency}} + \underbrace{\sum_{i\in\mathcal{N}}\tilde{\lambda}_i d\log A_i}_{\Delta\text{ Technology (Residual)}}.$$

This approach implements growth accounting but overlooks multiproduct firms engaged in joint production. Figure 5a illustrates the decomposition of cumulative changes in aggregate TFP growth from 2016 to 2022 under this assumption. Allocative efficiency term (in red) declined over this period. This outcome suggests that high-markup firms contracted further, resulting in a negative reallocation effect; however, the contribution of allocative efficiency exceeds that of the technology (residual) component, particularly during the COVID-19 pandemic and the subsequent high inflation period. To rationalize this disparity, the technology term, measured as a residual, must have increased by 5.6%.

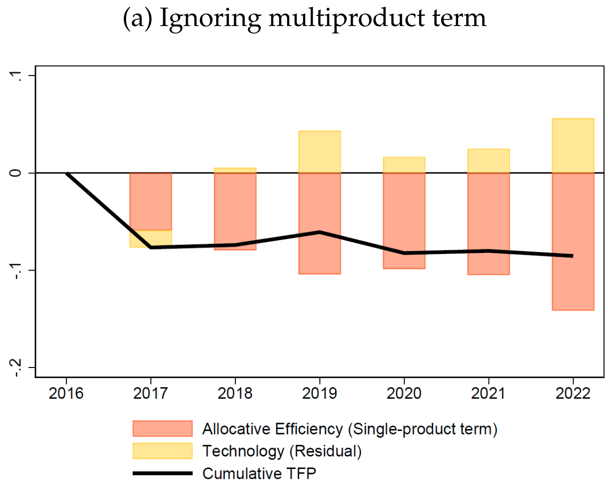
Next, we incorporate the multiproduct term using Proposition 3:

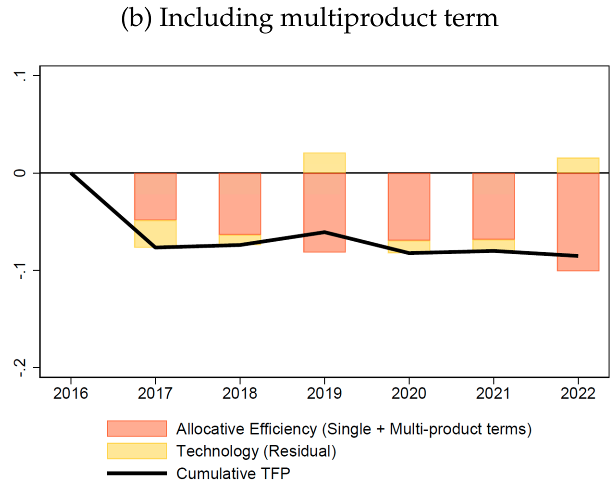
$$d\log TFP = \underbrace{\underbrace{\sum_{i\in\mathcal{N}}\tilde{\lambda}_i \mathrm{Cov}_{s_i}\left(d\log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right)}_{\text{Multiproduct term}} - \underbrace{\sum_{i\in\mathcal{N}}\tilde{\lambda}_i d\log\mu_i}_{\text{Firm-level Markup}} - \underbrace{\sum_{f\in\mathcal{F}}\tilde{\Lambda}_f d\log\Lambda_f}_{\text{Aggregate Factor Shares}}}_{\Delta\text{ Allocative Efficiency}} + \underbrace{\sum_{i\in\mathcal{N}}\tilde{\lambda}_i d\log A_i}_{\Delta\text{ Technology (Residual)}}.$$

Figure 5b presents the results incorporating the multiproduct term. Accounting for joint production reduces the measured allocative-efficiency decline from 14.1% under the single-product benchmark to 10.1%; equivalently, ignoring joint production overstates the deterioration in allocative efficiency by 40%. The implied technological residual is correspondingly small, rising by 1.6% rather than the 5.6% implied by the single-product benchmark. This result suggests that considering joint production considerably decreases the reallocation effects implied under the traditional assumption that firms produce only one product.

[18] CNEP (2023)

Figure 5: Cumulative TFP growth decomposition

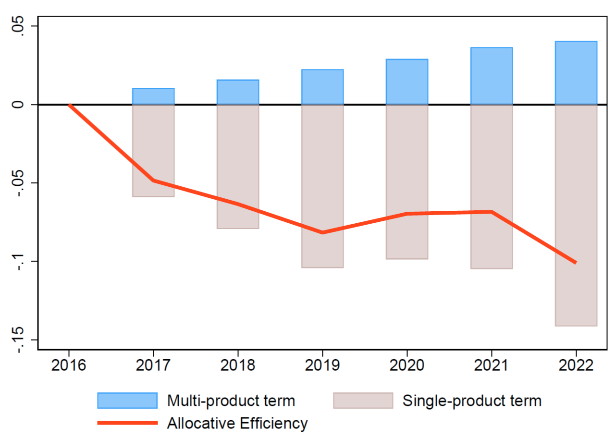




Notes: Panel (a) shows the cumulative change calculated by applying Corollary 1 repeatedly each year while Panel (b) does it applying Proposition 3. Technology (residual) is calculated by subtracting allocative efficiency from TFP growth in both panels.

Figure 6 decomposes the allocative efficiency in Figure 5b into multiproduct and single product terms. While the single product term dominates over the period, the multiproduct term is quantitatively important: ignoring it generates the 40% overestimation of allocative efficiency shown in Figure 5b.

Figure 6: Allocative efficiency components: Single vs multiproduct term



Notes: The Figure decomposes the cumulative change in allocative efficiency in Figure 5b single-product and multi-product terms.

This finding is consistent with the joint production mechanism described in Section 3. When firms engage in joint production, they create multiple products using common inputs. When a given product receives a shock, if firms face technological constraints to adjust their product mix (non-infinite elasticity of transformation as in the oil refinery example), firms will struggle to reallocate productive resources from one product to another. The reallocation through substitution among products within multi-product firms is attenuated, and reallocation is not materialized to the extent suggested under the single-product firm assumption.

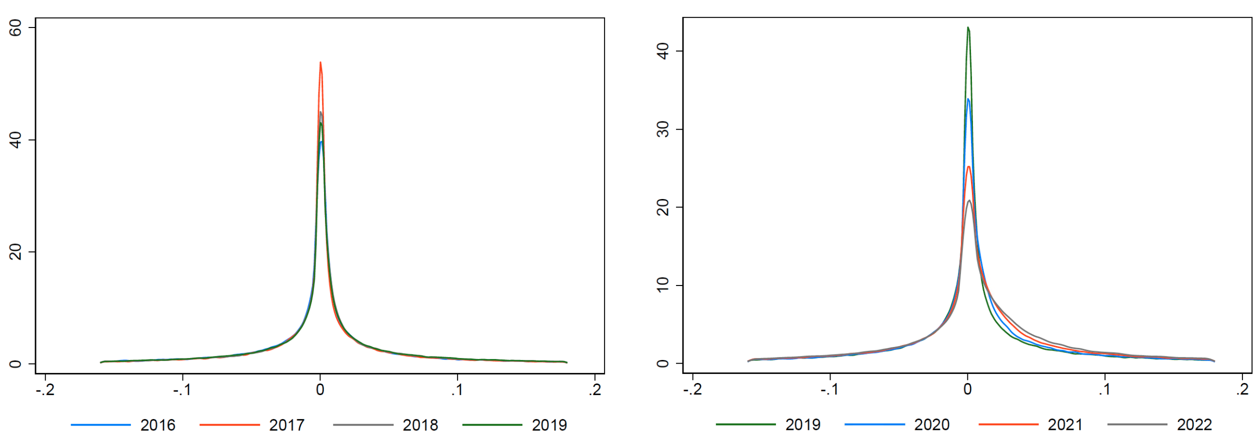
We confirm that this implication is robust to how own markups are measured. Appendix E reproduces the decomposition (Figure A3a) with firm-level accounting markups in place of the estimated product-level markups. The multiproduct term is slightly smaller, but qualitatively the result is unchanged: it continues to offset the single-product term, attenuating the measured allocative-efficiency decline relative to the single-product benchmark.

Finally, the granularity of the data allows us to track the distributional changes of joint production (the multiproduct term) that limit the extent of resource reallocation. Since the covariance degenerates to zero under the single-product firm assumption, the dispersion of covariance implies that joint-production forces are active. These distributions vary from period to period. Figure 7a plots the distribution for pre-COVID-19 (2016–2019), which is symmetric around 0, with slight differences from year to year.

Figure 7b presents the distribution after the onset of COVID-19, showing a shift to the right from year to year, resulting in a right-skewed distribution. This result suggests that the increase in the contribution from joint-production forces (the multiproduct term) was not driven by a few specific firms.

Figure 7: $\mathrm{Cov}_{s_i}\left(d\log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right)$ distributions by year

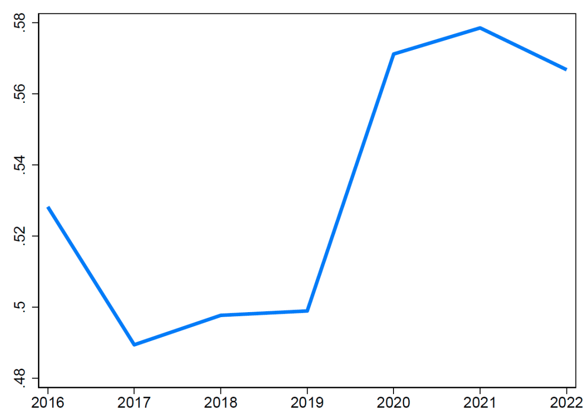
(a) 2016 - 2019 (b) 2019 - 2022



Notes: These Figures plot the distribution of firm level $\mathrm{Cov}_{s_i}\left(d\log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right)$ for each year.

Finally, Figure 8 plots the median variance of product-specific production changes across firms from 2016 to 2022. This figure provides suggestive evidence that aligns with the changing distribution of multi-product firms shown in Figure 7b and corresponds to the period of large contribution from the multi-product term in our decomposition. The increasing variance, particularly the sharp rise from 2019 to 2020 and its sustained high level thereafter, indicates that firms have been under greater pressure to adjust their product mix. This trend coincides with the timeframe when we observe the most substantial impact of the multi-product term on allocative efficiency. The temporal consistency between the increased variance in product-specific production changes and the heightened contribution of the multi-product term reinforces our model's emphasis on the importance of multi-product firms engaged in joint production, especially during major economic shocks like the COVID-19 pandemic.

Figure 8: Large product mix adjustments: suggestive evidence



Notes: This figure depicts the evolution of the median variance of product quantity changes, denoted as $Var_{\lambda_i}\left(d\log q_{ig}\right)$, from 2016 to 2022.

6 Ex-Ante Structural Results

This section develops a structural framework to predict how economies with multiproduct firms respond to shocks. While our previous analysis relied on observed price and factor share changes, we now model these endogenous responses explicitly. This structural framework allows us to move beyond ex-post measurement to ex-ante prediction of counterfactual scenarios. The framework complements our earlier results.

We show how to apply this framework to study the distance to the Pareto-efficient frontier when firms use joint production technology. This method compares output in an efficient equilibrium (with all markup wedges removed) to that in a distorted decentralized economy. Our analysis demonstrates how the theoretical results of previous studies, such as those of Hsieh and Klenow (2009) and Baqaee and Farhi (2020), change when firms engage in joint production. Since an economy without markups is unobservable, a model is necessary to analyze this counterfactual case.

6.1 The Nested CET-CES Model

We propose the nested CET-CES model, which provides a tractable framework for our analysis. We use setup of subsection 2.3 but impose the CET-CES functional form to the joint production function to the equation 6. We then derive a linear system for price and sales responses, allowing us to characterize the economy's response to shocks. The production technology is given by:

$$\underbrace{\left(\sum_{g\in\mathcal{G}}\delta_{ig}\left[q_{ig}\right]^{\frac{\sigma_i+1}{\sigma_i}}\right)^{\frac{\sigma_i}{\sigma_i+1}}}_{\text{Output bundle}} = \underbrace{A_i\left(\sum_{j\in\mathcal{N},g'\in\mathcal{G}} q_{i,jg'}^{\frac{\theta_i-1}{\theta_i}} + \omega_{i,L}L_i^{\frac{\theta_i-1}{\theta_i}} + \omega_{i,K}K_i^{\frac{\theta_i-1}{\theta_i}}\right)^{\frac{\theta_i}{\theta_i-1}}}_{\text{Input Bundle}}. \tag{12}$$

Here, σ_i represents the elasticity of transformation between different outputs, A_i denotes the productivity of firm i, and δ_{ig} are the output share parameters. The input bundle comprises intermediate inputs $q_{i,jg'}$, labor L_i, and capital K_i, aggregated using a CES function with an elasticity of substitution θ_i. Note that this class of models is highly general, nesting the nested CES system widely used in macroeconomics and international economics as a special case.[19] For single-output firms, the production function degenerates to:

$$q_i = \underbrace{A_i\left(\sum_{j\in\mathcal{N},g'\in\mathcal{G}} \omega_{i,jg'}q_{i,jg'}^{\frac{\theta_i-1}{\theta_i}} + \omega_{i,L}L_i^{\frac{\theta_i-1}{\theta_i}} + \omega_{i,K}K_i^{\frac{\theta_i-1}{\theta_i}}\right)^{\frac{\theta_i}{\theta_i-1}}}_{\text{Input bundle}}. \tag{13}$$

Furthermore, we specify the household utility function as a CES aggregator over final consumption goods. Formally, the representative household's utility function is given by:

$$U\left(c_{11},...,c_{ig},...,c_{NG}\right) = \left(\sum_{i\in\mathcal{N},g\in\mathcal{G}} \psi_{ig}c_{ig}^{\frac{\theta_0-1}{\theta_0}}\right)^{\frac{\theta_0}{\theta_0-1}} \tag{14}$$

where c_{ig} represents the consumption of product g from firm i, ψ_{ig} represents the taste parameter for each product, and θ_0 is the elasticity of substitution between products.

For analytical simplicity, we assume a uniform substitution elasticity within the firm's CES structure, though extending the model to incorporate further nesting would be straightforward.

[19]When multi-product firms are assumed to be a collection of single-product firms, we can model situations with different input ratios by assuming equation (13) for each product and treating that group as a firm, provided the necessary data is available.

6.2 Linear System for Price and Sales Responses

Using this model, we construct a system for solving the first-order response to primitive shocks ($\boldsymbol{A}, \boldsymbol{\mu}$) to the endogenous variables. This system of linear equations, derived from the model's first-order conditions, enables us to generate ex-ante predictions of how the economy will respond to counterfactual shocks. We begin with the multi-product firm's forward equation under the CET output function:

Proposition 5 (Multi-Product Firm's Forward Equation under CET Output Function)**.**

$$d\log p_{ig} = \underbrace{\sum_{j\in\mathcal{N},g'\in\mathcal{G}} \tilde{\Psi}_{ig,jg'}\left(d\log\mu_{jg'} - d\log A_j\right) + \sum_{f\in\mathcal{F}} \tilde{\Psi}_{ig,f} d\log\Lambda_f}_{\text{Indirect cost exposure}}$$
$$+ \underbrace{\sum_{j\in\mathcal{N},g'\in\mathcal{G}} c_j^R \tilde{\Psi}_{ig,jg'} d\log\Theta_{jg'}}_{\text{Indirect exposure to the Product Mix adjustment}},$$

where

$$d\log\Theta_{jg'} = \left(\frac{\sigma_j}{\sigma_j+1}\right) d\log\mu_{jg'}/\mu_{jr} + \frac{1}{\left(1+\sigma_j\right)}\left[d\log\lambda_{jg'}/\lambda_{jr}\right],$$

and $c_j^R = \frac{mc_{jr}q_{jr}}{\sum mc_{jg}q_{jg}}$ *is a reference product cost share. Here, r denotes a reference good.*

This equation describes how changes in unit prices within a firm due to markups, productivity shocks, and price changes associated with endogenous product mix adjustments are transmitted through production networks to other firms' products. The first term illustrates the effect of exposure to common cost shocks on prices, a force present in standard production network models. The second term, unique to the joint production model, indicates the exposure of firms with nonlinear production possibility frontiers to modify their product mix due to reallocation, which affects endogenous unit costs. The magnitude of this effect depends on the transformation elasticity, with lower elasticities leading to larger cost effects. As σ approaches infinity, the cost of product-mix adjustment vanishes, as firms can freely adjust their product mix.

Next, we consider backward propagation:

Proposition 6 (Backward Propagation)**.**

$$\lambda_{ig} d\log\lambda_{ig} = -\sum_{j\in\mathcal{N},g'\in\mathcal{G}} \lambda_{jg'}\left(\Psi_{jg',ig} - \mathbf{1}\left(ig = jg'\right)\right) d\log\mu_{jg'}$$
$$+ \sum_{k\in\mathcal{N},g''\in\mathcal{G}} \frac{\lambda_{kg''}}{\mu_{kg''}}\left(1-\theta^k\right)\mathrm{Cov}_{\tilde{\Omega}_{(kg'',:)}}\left(d\log p, \Psi_{(:,ig)}\right).$$

This equation, based Baqaee and Farhi (2020) methodology, describes the sales and factor share response. Notably, it does not contain σ, indicating that joint production does not directly enter the changes in sales share via substitution effects. The equation shows how shocks propagate through the production network, affecting markups and quantities of each product via changes in upstream suppliers' price indices and productivity.

By combining the forward equation from Proposition 5 and the backward equation from Proposition 6, we obtain a complete system of linear equations that characterizes the economy's response to shocks. This system consists of $2 \times (1 + \mathcal{N} \times \mathcal{G} + \mathcal{F})$ equations and $2 \times (1 + \mathcal{N} \times \mathcal{G} + \mathcal{F})$ unknowns, where $\mathcal{N}$ is the number of firms, $\mathcal{G}$ is the number of products, and $\mathcal{F}$ is the number of factors. This system of equations fully characterizes the first-order response of all endogenous variables to any combination of productivity or markup shocks. By solving this linear system using standard matrix algebra, we can conduct counterfactual analyses and evaluate the impact of various shocks on the economy's equilibrium outcomes.

6.3 Distance to the Pareto-Efficient Frontier

Using our model, we can characterize the distance to the Pareto-efficient frontier when introducing distortions, allowing us to predict efficiency losses from counterfactual changes in markups or other distortions. Let $\mathcal{L} = -\Delta \log TFP$ denote the welfare loss.

Proposition 7 (Distance to the Pareto-Efficient Frontier)**.** *Under joint production, starting at an efficient equilibrium, up to second order, and in response to the introduction of distortions, the change in TFP is given by Domar-weighted Harberger triangles:*

$$\Delta \log TFP = \frac{1}{2} \sum_{ig} \lambda_{ig} d \log q_{ig} d \log \mu_{ig}.$$

Equivalently, since $\mathcal{L} = -\Delta \log TFP$*, the welfare loss is*

$$\mathcal{L} = -\frac{1}{2} \sum_{ig} \lambda_{ig} d \log q_{ig} d \log \mu_{ig}.$$

Here λ_{ig} *is the Domar weight of product* g *from firm* i*,* q_{ig} *is the quantity, and* μ_{ig} *is the markup.*

This result shows that the second-order TFP change from introducing distortions is determined by three statistics: the Domar weight of each product, the magnitude of the wedge, and the induced change in product quantity. The quantity change can be derived from sales and price changes given by our linear system, using the relationship $d \log q = d \log \lambda - d \log p$. Since this system includes the transformation elasticity σ, the welfare loss generally depends on the value of σ.

To illustrate how joint production affects the distance to the frontier, we provide analytical solutions for a special case.

6.3.1 Horizontal Economy with Joint Production

We consider a horizontal economy similar to Hsieh and Klenow (2009) but we use joint production technologies. This allows us to investigate within-firm markup heterogeneity in the presence of production transformation constraints. In an economy with a representative consumer (CES utility with elasticity θ), $\mathcal{N}$ firms each use a shared input L to produce $\mathcal{G}$ products using CET technology (elasticity of transformation σ). Markups μ_{ig} are heterogeneous across products and firms.

Proposition 8 (The Distance to the Frontier in the Horizontal Economy)**.** *Consider a horizontal economy with a representative consumer with CES utility with elasticity θ. Each firm i uses a shared input L to produce multiple products using CET technology with elasticity of transformation σ. Markups μ_{ig} are heterogeneous across products and firms. Starting from an efficient equilibrium, the second-order distance to the Pareto-efficient frontier from introducing markup distortions is*

$$\mathcal{L} = \frac{1}{2}\theta\left[\mathrm{Var}_{\bar{\lambda}}\left(d\log\mu_i\right) + \frac{\sigma}{\sigma+\theta}\mathbb{E}_{\bar{\lambda}}\left\{\mathrm{Var}_{\boldsymbol{s}_i}\left(d\log\mu_{ig}\right)\right\}\right].$$

Equivalently, by the law of total variance,

$$\mathcal{L} = \frac{1}{2}\theta\left[\mathrm{Var}_{\lambda}\left(d\log\mu_{ig}\right) - \frac{\theta}{\sigma+\theta}\mathbb{E}_{\bar{\lambda}}\left\{\mathrm{Var}_{\boldsymbol{s}_i}\left(d\log\mu_{ig}\right)\right\}\right],$$

where

$$\lambda_i = \sum_g \lambda_{ig}, \qquad s_{ig} = \frac{\lambda_{ig}}{\lambda_i}, \qquad d\log\mu_i = \mathbb{E}_{\boldsymbol{s}_i}\left[d\log\mu_{ig}\right],$$

$\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \ldots, \lambda_{NG})$ *is the product-level Domar-weight vector, and* $\bar{\boldsymbol{\lambda}} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ *is the firm-level Domar-weight vector.*

Proposition 8 characterizes the distance to the frontier in a horizontal economy with joint production and heterogeneous markups. The distance to the frontier comprises two markup variances. The first term, the between-firm dispersion in average markup changes $\mathrm{Var}_{\bar{\lambda}}(d\log\mu_i)$, captures the misallocation that would arise even if each firm produced a single product. The second term, which is related to joint production, means that the finite elasticity of transformation σ attenuates the contribution of within-firm markup dispersion. The fraction of within-firm dispersion that translates into welfare losses is $\sigma/(\sigma+\theta)$, so joint production attenuates this component by $\theta/(\sigma+\theta)$ relative to the separable product-line benchmark. As σ increases and approaches infinity, the force of attenuation associated with joint production approaches zero. Conversely, as σ

approaches zero (which implies Leontief production technology), the importance of within-firm markup dispersion decreases. We give this result in the following formal corollary.

Corollary 3 (Limit Cases)**.** *The distance to the frontier simplifies in extreme cases of the elasticity of transformation:*

(i) As $\sigma \to \infty$ (perfect transformability between products):

$$\mathcal{L} = \frac{1}{2}\theta \operatorname{Var}_{\lambda}\left(d \log \mu_{ig}\right).$$

(ii) As $\sigma \to 0$ (Leontief product mix):

$$\mathcal{L} = \frac{1}{2}\theta \operatorname{Var}_{\bar{\lambda}}\left(\mathbb{E}_{s_i}\left[d \log \mu_{ig}\right]\right).$$

In the case of perfect transformability, misallocation depends on the variance of markups across all products. This term can be obtained by treating each product as an independent firm and applying the results of Baqaee and Farhi (2020). Conversely, in the Leontief case, firms cannot adjust their product mix at all, so only the dispersion in firm-average markup changes is relevant.

These results imply that, in the absence of within-firm markup dispersion, the term related to joint production disappears regardless of the value of σ. However, this reasoning is specific to horizontal economies. This relationship easily breaks down in more complex economic structures that include firm-to-firm networks, and σ remains essential even when markups within firms are homogeneous, because product-mix constraints interact with upstream and downstream propagation.

6.4 Application to Chile

As a complementary, ex-ante illustration, we apply the horizontal economy of Proposition 8 to quantify efficiency losses from markup distortions in Chile in 2018. This setting yields transparent, closed-form sufficient statistics that cleanly isolate the role of joint production; we therefore read it as a gauge of the mechanism's empirical relevance rather than as a structural estimate of aggregate misallocation in Chile. To implement Proposition 8, we require estimates of both the CET transformation elasticity (σ) and the substitution elasticity (θ). Following Arkolakis et al. (2023), we set $\theta = 2.5$ as our benchmark and report a Cobb-Douglas sensitivity with $\theta = 1$.

For the transformation elasticity, we use $\sigma = 1.589$, our estimate from Section 3, which is also the value used in our product markup estimation. We compare the value of the distance to the frontier with the case of independent product lines ($\sigma \to \infty$), which, under constant returns to scale, is equivalent to assuming single-product firms.

For markups, we employ the product-level estimates constructed in Section 4, which are consistent with our CET specification and $\sigma = 1.589$. Our primary focus is on comparing economies with joint production to cases with independent product lines.

Table 2: The Distance to the Frontier

Specification	σ	θ	Distance to Frontier
Benchmark economy	1.589	2.5	20.05%
Independent product lines	$\to \infty$	2.5	26.84%
Cobb-Douglas economy	1.589	1.0	9.02%
Cobb-Douglas independent product lines	$\to \infty$	1.0	10.73%

Notes: The table reports welfare losses (Harberger triangles) as a percentage of GDP under different assumptions on the elasticity of transformation (σ) and substitution (θ). The independent product lines case ($\sigma \to \infty$) corresponds to the standard single-product firm assumption in the literature. The Cobb-Douglas row isolates the role of demand elasticity by setting $\theta = 1$.

As reported in Table 2, our benchmark economy with joint production ($\sigma = 1.589$, $\theta = 2.5$) implies welfare losses of 20.05% of GDP. Under the independent-product-line benchmark ($\sigma \to \infty$), which shuts down joint-production attenuation and is equivalent to treating each product as a separate single-product firm, the same wedges imply losses of 26.84%. Joint production therefore attenuates the measured welfare loss by about 25% relative to this benchmark.

This pattern reflects the central mechanism of the paper: joint production constrains within-firm reallocation, so separable technology overstates the reallocation a given set of wedges can induce, and hence the implied welfare loss. The demand elasticity sharpens the comparison. A higher θ raises the product-level reallocation that markup dispersion would call for under the independent-product-line benchmark, while joint production limits firms' capacity to accommodate it, so the benchmark's overstatement grows with θ. Formally, joint production removes a fraction $\theta/(\sigma + \theta)$ of the within-firm markup-dispersion term, which is increasing in θ for a given σ. The Cobb-Douglas case ($\theta = 1$) confirms this: losses fall to 9.02% under joint production and 10.73% under independent product lines, an attenuation of about 16%, well below the roughly 25% obtained under CES preferences, where demand is more elastic and losses are larger in level as well.

This application illustrates the framework and underscores the importance of accounting for joint production when measuring misallocation in economies with multiproduct firms. By treating multiproduct firms as collections of independently adjustable product lines, the single-product benchmark overstates the efficiency gains available from removing markup distortions, by roughly one quarter under our benchmark calibration.

7 Conclusion

This paper develops a theoretical framework for aggregating distortions in production networks with multiproduct firms. We show that when firms jointly produce multiple products using common inputs, standard misallocation accounting can overstate the extent of resource reallocation implied by observed wedges. The reason is that joint production constrains firms' ability to adjust their product mix: products with high cumulative wedges may remain underproduced not only because of distortions, but also because technological constraints limit within-firm reallocation.

We derive sufficient statistics that decompose aggregate TFP changes into technology and allocative-efficiency components in production networks with joint production. The key additional term is a firm-level covariance between product-level price changes and cumulative wedges. This term captures how technological constraints within multiproduct firms attenuate the reallocation effects that would arise under the standard single-product-firm assumption.

We implement the framework using comprehensive Chilean firm-to-firm transaction data to assess the quantitative relevance of this mechanism. Applying the framework to a period of stagnant TFP growth in Chile, we find that a standard single-product benchmark attributes a 14.1% decline to allocative efficiency. Accounting for joint production reduces this measured decline to 10.1%. Thus, ignoring joint production overestimates the decline in allocative efficiency by 40% relative to the joint-production estimate.

The results show that joint production is not only a theoretical possibility but also a quantitatively relevant force for misallocation accounting. Multiproduct firms do not behave as collections of independent product lines: demand shocks to one product affect the production and pricing of other products within the same firm. These within-firm technological constraints attenuate the amount of reallocation that single-product models would infer from observed wedges.

References

Acevedo, P., Luttini, E., Pizarro, M., and Quevedo, D. (2023). Invoices rather than surveys: using ML to nominal and real indices.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica: journal of the Econometric Society*, 83(6):2411–2451.

Almunia, M., Antràs, P., Lopez-Rodriguez, D., and Morales, E. (2021). Venting out: Exports during a domestic slump. *The American economic review*, 111(11):3611–3662.

Arkolakis, C., Huneeus, F., and Miyauchi, Y. (2023). Spatial production networks. (30954).

Bai, X., Fernández-Villaverde, J., Li, Y., and Zanetti, F. (2024). The causal effects of global supply chain disruptions on macroeconomic outcomes: Evidence and theory. *SSRN Electronic Journal*.

Baqaee, D., Burstein, A., Duprez, C., and Farhi, E. (2023). *Supplier Churn and Growth: A Micro-to-Macro Analysis*. National Bureau of Economic Research.

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The quarterly journal of economics*, 135(1):105–163.

Basu, S. and Fernald, J. G. (2002). Aggregate productivity and aggregate technology. *European economic review*, 46(6):963–991.

Bernard, A. B., Redding, S. J., and Schott, P. K. (2010). Multiple-product firms and product switching. *The American economic review*, 100(1):70–97.

Bernard, A. B., Redding, S. J., and Schott, P. K. (2011). Multiproduct firms and trade liberalization. *The Quarterly Journal of Economics*, 126(3):1271–1318.

Bigio, S. and La'O, J. (2020). Distortions in production networks*. *The quarterly journal of economics*, 135(4):2187–2253.

Boehm, C. E., Flaaen, A., and Pandalai-Nayar, N. (2019). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake. *The review of economics and statistics*, 101(1):60–75.

Boehm, J. and Oberfield, E. (2023). Growth and the fragmentation of production.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021a). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of monetary economics*, 121:1–14.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021b). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of monetary economics*, 121:1–14.

Burstein, A., Cravino, J., and Rojas, M. (2024). Input price dispersion across buyers and misallocation.

Cairncross, J. and Morrow, P. (2023). Multi-product markups.

Carrillo, P., Donaldson, D., Pomeranz, D., and Singhal, M. (2023). *Misallocation in Firm Production: A Nonparametric Analysis Using Procurement Lotteries*. National Bureau of Economic Research.

Carvalho, V. M., Nirei, M., Saito, Y. U., and Tahbaz-Salehi, A. (2020). Supply chain disruptions: Evidence from the great east japan earthquake*. *The quarterly journal of economics*, 136(2):1255–1321.

CNEP, C. N. E. . P. C. (2023). Informe anual de productividad 2023.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica: journal of the Econometric Society*, 84(2):445–510.

Dhyne, E., Petrin, A., Smeets, V., and Warzynski, F. (2017). Multi product firms, import competition, and the evolution of firm-product technical efficiencies. (23637).

Dhyne, E., Petrin, A., Smeets, V., and Warzynski, F. (2022). Theory for extending single-product production function estimation to multi-product settings. (30784).

Diewert, W. E. (1971). An application of the shephard duality theorem: A generalized leontief production function. *The journal of political economy*, 79(3):481–507.

Diewert, W. E. (1973). Functional forms for profit and transformation functions. *Journal of Economic Theory*.

Ding, X. (2023). *Industry Linkages from Joint Production*. U.S. Census Bureau, Center for Economic Studies.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How costly are markups? *Journal of Political Economy*, 131(7):1619–1675.

Fujiy, B. C., Ghose, D., and Khanna, G. (2022). Production networks and firm-level elasticities of substitution.

Giroud, X. and Mueller, H. M. (2019). Firms' internal networks and local economic shocks. *The American economic review*, 109(10):3617–3649.

Hall, R. (1988). The relation between price and marginal cost in U.S. industry. *The journal of political economy*, 96(5):921–947.

Hall, R. E. (1973). The specification of technology with several kinds of output. *The journal of political economy*, 81(4):878–892.

Hottman, C. J., Redding, S. J., and Weinstein, D. E. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in china and india. *The quarterly journal of economics*, 124(4):1403–1448.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of economic studies*, 45(3):511–518.

Kikkawa, A. K. (2022). Imperfect competition in firm-to-firm trade. *Journal of the European Economic Association*, 20(5):1933–1970.

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking*, 27(4):1241–1277. Part 2.

Klenow, P. J. and Willis, J. L. (2016). Real rigidities and nominal price changes. *Economica*, 83(331):443–472.

Klette, T. J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *The journal of political economy*, 112(5):986–1018.

Lau, L. J. (1972). Profit functions of technologies with multiple inputs and outputs. *The review of economics and statistics*, 54(3):281–289.

Matsuyama, K. (2025). Homothetic non-CES demand systems with applications to monopolistic competition. *Annual Review of Economics*, 17:261–292.

Mayer, T., Melitz, M. J., and Ottaviano, G. I. P. (2014). Market size, competition, and the product mix of exporters. *The American economic review*, 104(2):495–536.

Mayer, T., Melitz, M. J., and Ottaviano, G. I. P. (2021). Product mix and firm productivity responses to trade competition. *The review of economics and statistics*, 103(5):874–891.

Petrin, A. and Levinsohn, J. (2012). Measuring aggregate productivity growth using plant-level data. *The Rand journal of economics*, 43(4):705–725.

Powell, A. A. and Gruen, F. H. G. (1968). The constant elasticity of transformation production frontier and linear supply system. *International economic review*, 9(3):315–328.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of economic dynamics*, 11(4):707–720.

Solow, R. M. (1957). Technical change and the aggregate production function.

Valmari, N. (2023). Estimating production functions of multiproduct firms. *The Review of economic studies*.

Wang, W. and Yang, E. (2023). Multi-product firms and misallocation. *Journal of development economics*, 163(103102):103102.

Appendix for "Aggregating Distortions in Networks with Multi-Product Firms"

A Additional Figures and Tables

A.1 Additional Empirical Results for Reduced-Form Evidence

In this appendix section, we present additional Figures and Table for the event study analysis shown in Section 3.

Validation of COVID-19 Lockdowns as Demand Shocks

This appendix validates our use of Chilean COVID-19 lockdowns in March 2020 as demand shocks to intermediate input transactions. We demonstrate empirically, using firm-level transaction data, that these lockdowns led to substantial reductions in intermediate input purchases.

We posit that intermediate input transactions declined between suppliers in unaffected (gray) counties and buyers in counties that experienced early COVID-19 lockdowns (red). To test this hypothesis, we estimate the following reduced-form specification at the buyer level:

$$\log M_{it} = \beta \, Lockdown_{it} + FE_t + FE_i + \varepsilon_{it}, \tag{15}$$

where M_{it} denotes total intermediate input purchases of a firm i at time t and $Lockdown_{it}$ is a dummy variable equal to one if firm i's location was under lockdown at time t, and is zero otherwise. To address potential bias arising from buyers in lockdown areas who purchased from suppliers in lockdown areas, we restricted the sample by including only buyers with suppliers in non-lockdown areas.

Table A1: Lockdown and intermediate input purchases

	(1)	(2)	(3)
Lockdown Dummy	-0.222***	-0.230***	-0.191***
	(0.0524)	(0.00521)	(0.0589)
Firm FE	Y	Y	Y
Time FE	N	N	Y
Sector × Time FE	N	Y	N
Restricted sample	N	N	Y

Notes: The Table reports the results of estimating equation (15) by ordinary least squares (OLS), clustered at the firm-municipality level. The sample periods are January 2019 to March 2020. Columns (1) and (2) report results for the full sample. Column (3) presents the results restricted to firms with no suppliers in the lockdown area. Three stars indicate statistical significance at the 1% level.

The results confirm our hypothesis: The coefficient of interest, β, is negative, indicating that purchases of intermediate inputs from lockdown counties decreased by about 20% on average. This result confirms that we can interpret the decrease in purchases as a negative demand shock to intermediate inputs sold by firms in non-lockdown regions to buyers in lockdown regions.

Characteristics of Treatment Firms

Table A2 displays the characteristics of treated and control firms.

Table A2: Characteristics of treatment firms

	Treatment Firms	Control Firms
Number of firms	1,432	155,247
Number of workers	3	3
Number of products sold	16	10
Number of providers	98	134
Number of buyers	24	8
Annual revenue (million pesos)	412	319
Annual total intermediate purchases (million pesos)	107	59
Share of firms in manufacturing	0.21	0.24
Share of firms in Retail and wholesale	0.44	0.39
Share of firms in Services	0.22	0.21

Notes: This Table presents the characteristics of treated firms (those whose major product buyers experienced lockdowns in March 2020) and control firms, showing values from February 2020, the month before the shock. The rows for the number of workers, products sold, providers, buyers, revenue, and total intermediate purchases display the median of each statistic. The industry shares indicate the proportion of firms within each group that belong to specific industries.

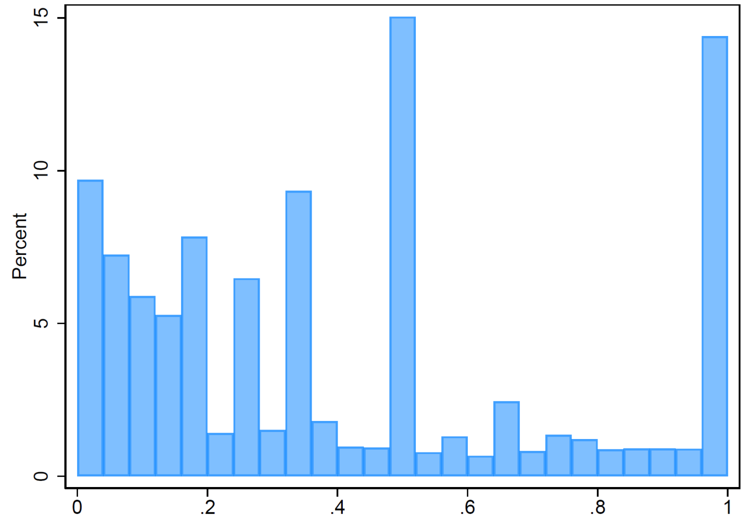
Firms Sell Different Products to Distinct Sets of Buyers

We construct the following measure to characterize the heterogeneity from the intermediate inputs buyer perspective across products and within firms:

$$S_i = \frac{\text{number of buyers of the main product of firm } i}{\text{number of buyers of firm } i},$$

where the main product is the one that has the largest sales within firm i in 2018. Figure A1 presents the distribution of this measure across firms.

Figure A1: Buyer heterogeneity across product within firm



Notes: Histogram of the number of buyers buying the main product of firm i / number of buyers in firm i for a multiproduct firm. The main product is the product with the highest sales within that firm. Data are from 2018.

If the buyers of the seller's main product and its other products were the same, S_i would be one. Some mass exists at $S_i = 1$ but for more than 50% of multiproduct firms; however, buyers of the main product constitute less than 50% of the total buyer-firms base. The fact that each product has a distinct set of buyers ensures that we can construct a sample where the main product experiences a demand shock while the other products do not.

DiD Specification

This section presents sensitivity results using the Difference-in-Differences (DiD) specification, showing responses in both quantity and price.

$$\log X_{igt} = \beta \underbrace{Lockdown_{i,t}}_{\text{shock to main product}} + FE_{ig} + FE_t + \varepsilon_{igt},$$

where $Lockdown_{i,t} = 1$ if firm i 's main product experienced the demand shock and after March 2020.

Columns (1) and (2) correspond to the Figure 4 of the main text. Columns (3) and (4) include the input price index as a control variable. Columns (5) and (6) restrict the sample to large firms. Columns (7) and (8) limit the analysis to manufacturing firms only. Columns (9) and (10) incorporate fixed effects that vary over time at the level of the harmonized product code (CUP). Columns (11) and (12) replace the binary treatment with a continuous measure using the share of transaction values to lockdown destinations in the firm's main product.

Table A3: Average treatment effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\log q$	-0.117***		-0.117***		-0.102***		-0.106***		-0.125***		-0.159***	
	(0.0046)		(0.0046)		(0.0053)		(0.0188)		(0.0049)		(0.0062)	
$\log p$		0.013***		0.013***		0.015***		0.014		0.012***		0.013***
		(0.0022)		(0.0022)		(0.0024)		(0.0096)		(0.0023)		(0.0024)
Input price control	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Large firms	N	N	N	N	Y	Y	N	N	N	N	N	N
Only manufacturing firms	N	N	N	N	N	N	Y	Y	N	N	N	N
Time × product FE	N	N	N	N	N	N	N	N	Y	Y	N	N
Continuous treatment	N	N	N	N	N	N	N	N	N	N	Y	Y
Observations	7,693,066	7,693,066	7,669,848	7,669,848	4,394,166	4,394,166	1,399,617	1,399,617	7,594,918	7,594,918	7,693,066	7,693,066

Notes: Input price control indicates inclusion of the Tornqvist input price index. Large firms restricts to firms above the 80th percentile in total sales. Only manufacturing firms restricts to the manufacturing sector. Time × Product FE are time-varying fixed effects at the harmonized product level. Continuous treatment uses the share of transaction values to lockdown destinations in firm's main product instead of the binary lockdown variable. *** denote significance at 1%.

Distribution of the Number of Products

We use the 2018 data to describe the main features of the firm-to-firm trade patterns. Of all firms, 75% produce multiple products, and these firms account for 98.94% of intermediate input transaction value. Table A4 illustrates the distribution of products per firm, weighted by firm-to-firm transaction values.

Table A4: Distribution of product numbers

Percentile	**Number of products (Unweighted)**	**Number of products (Weighted by transaction value)**
1%	1	1
5%	1	2
10%	1	4
25%	2	36
50%	7	475
75%	26	2,459
90%	119	32,195
95%	290	37,422
99%	1,253	62,372

Notes: The Table presents the distribution of product numbers for 2018. The left column shows the number of products without weighting, while the right column displays the number of products weighted by the intermediate product transaction volumes of the firms.

Finally, we provide aggregate firm-level statistics for our growth accounting.

Table A5: Aggregate firm-level statistics

Year	Count	Sales	Wagebill	Employment
2016	110,451	262,506	40,260	4,242,555
2017	114,480	277,960	43,691	4,349,248
2018	115,916	330,486	44,688	4,349,454
2019	116,706	336,386	47,299	4,425,780
2020	102,306	310,317	44,053	3,935,883
2021	105,651	376,220	51,642	4,166,838
2022	105,032	454,818	59,148	4,266,972

Notes: Count stands by the number of firms when sales and wage bills are yearly aggregates expressed in millions of pesos. Employment represents the headcount of yearly workers included in the sample.

B A CET Cost Function and Lerner Index

We solve the firm's profit maximization problem in the setting at Section 3 and show that the optimal markup in joint production follows the standard Lerner formula.

Profit Function. The firm's profit is

$$\Pi = \sum_{i=1}^{N} \Big[p_i \, q_i \Big] - C(q_1, \ldots, q_N).$$

We want to solve for each p_i that maximizes Π, showing that the ratio $p_i / \big(\partial C / \partial q_i\big)$ is given by a standard Lerner markup formula.

We consider how p_i affects profit. Because q_i depends only on p_i, we have

$$\frac{\partial \Pi}{\partial p_i} = \frac{\partial}{\partial p_i} \Big[p_i \, q_i(p_i) \Big] \; - \; \frac{\partial}{\partial p_i} C(\ldots).$$

The cost $C(\ldots)$ depends on p_i only through q_i, so

$$\frac{\partial}{\partial p_i} C(\ldots) = \frac{\partial C}{\partial q_i} \frac{dq_i}{dp_i} \quad \text{(the partials w.r.t. } q_j \text{ for } j \neq i \text{ vanish, since } p_j \text{ does not enter } q_i\text{)}.$$

Hence,

$$\frac{\partial}{\partial p_i} \Big[p_i \, q_i \Big] = q_i + p_i \, \frac{dq_i}{dp_i}.$$

But from $q_i = D_i \, p_i^{-\theta_i}$,

$$\frac{dq_i}{dp_i} = -\theta_i \, D_i \, p_i^{-\theta_i - 1} = -\theta_i \, \frac{q_i}{p_i}.$$

Thus

$$\frac{\partial}{\partial p_i} \Big[p_i \, q_i \Big] = p_i \Big[-\theta_i \, \frac{q_i}{p_i} \Big] + q_i = (1 - \theta_i) \, q_i.$$

Meanwhile,

$$\frac{\partial}{\partial p_i} C(q_1, \ldots, q_N) = mc_i \, \frac{dq_i}{dp_i} = mc_i \Big[-\theta_i \, \frac{q_i}{p_i} \Big].$$

Therefore

$$\frac{\partial \Pi}{\partial p_i} = (1 - \theta_i) \, q_i \; - \; mc_i \Big[-\theta_i \, \tfrac{q_i}{p_i} \Big] = (1 - \theta_i) \, q_i + \theta_i \, \tfrac{q_i}{p_i} \, mc_i.$$

Setting $\frac{\partial \Pi}{\partial p_i} = 0$ gives

$$(1 - \theta_i) \, q_i + \theta_i \frac{q_i}{p_i} \, mc_i = 0.$$

Divide through by $q_i > 0$:

$$(1-\theta_i) + \theta_i \frac{mc_i}{p_i} = 0 \quad \Longrightarrow \quad \theta_i \frac{mc_i}{p_i} = \theta_i - 1 \quad \Longrightarrow \quad \frac{p_i}{mc_i} = \frac{\theta_i}{\theta_i - 1}.$$

So the optimal price for good i is

$$p_i = \frac{\theta_i}{\theta_i - 1}\, mc_i,$$

matching the usual Lerner Index ratio $\frac{\theta_i}{\theta_i - 1}$, even though the cost function couples all outputs $(q_1, \ldots, q_N)$.

C Monte Carlo Validation of the σ Estimator

The IV estimator of σ in Section 3.3 relies on the maintained assumption that relative markups μ_g/μ_m do not respond to the main-product demand shock $\Delta \log D_m$. A natural concern is that this assumption fails in a way that artificially generates a finite $\hat{\sigma}$ and a spurious rejection of product-line separability ($\sigma = \infty$). This appendix addresses that concern by formalizing the informal argument in footnote 17: under Marshall's Second Law of Demand, a standard property of variable-markup demand systems, the residual bias works in the *opposite* direction, biasing $\hat{\sigma}$ *upward*. The rejection of separability in Table 1 is therefore, if anything, conservative. We establish this analytically (Proposition 9), quantify the magnitude at roughly 18% under the Edmond et al. (2023) calibration of the Klenow and Willis (2016) Kimball aggregator, and confirm the analytical prediction by Monte Carlo simulation.

C.1 IV Decomposition

Let $Z \equiv \Delta \log D_m$ denote the structural demand shock to the main product, instrumented in the empirical implementation by main-product buyer lockdown exposure. Combining the relative-price relation in footnote 16 with the exclusion restriction $Z \perp \Delta \log(a_g/a_m)$, the population analog of the IV coefficient β in equation (11) decomposes as

$$\beta = \frac{1}{\sigma} + \underbrace{\frac{\mathrm{Cov}\big(Z, \Delta \log(\mu_g/\mu_m)\big)}{\mathrm{Cov}\big(Z, \Delta \log(q_g/q_m)\big)}}_{\text{markup-response bias}}. \tag{16}$$

Under the maintained assumption that relative markups do not respond to Z, the second term vanishes; the population estimand then recovers $1/\sigma$, and $\hat{\sigma} \equiv 1/\hat{\beta}$ is consistent for σ under standard regularity conditions. The remainder of this appendix asks how large the residual bias is when this assumption is relaxed in the direction favored by the variable-markup literature.

C.2 Bias under Marshall's Second Law

We relax the constant-elasticity demand specification in equation (9) to allow the demand elasticity to depend on relative demand, $\theta_i = \theta(r_i)$ with $r_i \equiv q_i/D_i$. Marshall's Second Law of Demand (MSL) is the property that $\theta(\cdot)$ is decreasing in r, so that the implied markup $\mu_i = \theta_i/(\theta_i - 1)$ is increasing in relative demand. MSL has become the standard restriction on demand families generating variable markups in the modern non-CES literature; see Matsuyama (2025) for a recent review, with widely used parametric implementations including Kimball (1995), Klenow and Willis (2016), and Edmond et al. (2023). Equivalently, the steady-state markup pass-through

$$\alpha \equiv \left. \frac{\partial \log \mu}{\partial \log r} \right|_{r=1} \geq 0,$$

where $\bar{\theta} \equiv \theta(1) > 1$ denotes the steady-state demand elasticity.

Proposition 9. *Suppose $\mu_i = \mu(r_i)$ with $\alpha \geq 0$. Then the markup-response term in (16) is nonpositive, and*

$$\beta = \frac{1}{\sigma} \cdot \frac{1}{1 + \bar{\theta}\alpha} \leq \frac{1}{\sigma}, \qquad \hat{\sigma} \geq \sigma, \tag{17}$$

with strict inequality whenever $\alpha > 0$.

Proof. Working to first order, the equilibrium responses are additively separable in Z and the relative productivity shock $\Delta \log(a_g/a_m)$. Under $Z \perp \Delta \log(a_g/a_m)$, the productivity shock contributes to neither covariance in (16), so we may set $\Delta \log a_g = \Delta \log a_m = 0$ when deriving the bias term. Log-linearizing around the symmetric steady state $q_m = q_g = D_m = D_g = 1$, the first-order system consists of demand, $\Delta \log q_i = \Delta \log D_i - \bar{\theta}\, \Delta \log p_i$; markup pricing, $\Delta \log p_i = \Delta \log \mu_i + \Delta \log mc_i$; the MSL pass-through, $\Delta \log \mu_i = \alpha\,(\Delta \log q_i - \Delta \log D_i)$; and the CET relative marginal cost implied by equation (8), $\Delta \log mc_g - \Delta \log mc_m = \sigma^{-1}\, \Delta \log(q_g/q_m)$. Imposing $\Delta \log D_m = Z$ and $\Delta \log D_g = 0$ and solving for the relative quantities and markups gives

$$\Delta \log \frac{q_g}{q_m} = -\frac{\sigma(1 + \bar{\theta}\alpha)}{\sigma + \bar{\theta}(1 + \sigma\alpha)}\, Z, \qquad \Delta \log \frac{\mu_g}{\mu_m} = \frac{\bar{\theta}\alpha}{\sigma + \bar{\theta}(1 + \sigma\alpha)}\, Z.$$

Both are linear in Z, so the covariance ratio in (16) equals the slope ratio,

$$\frac{\operatorname{Cov}\big(Z, \Delta \log(\mu_g/\mu_m)\big)}{\operatorname{Cov}\big(Z, \Delta \log(q_g/q_m)\big)} = -\frac{\bar{\theta}\alpha}{\sigma(1 + \bar{\theta}\alpha)} \leq 0,$$

strictly so when $\alpha > 0$. Substituting into (16) yields (17). □

Proposition 9 signs the bias without any parametric restriction on $\mu(\cdot)$: under any MSL-consistent demand system, $\hat{\sigma}$ overstates σ, with the magnitude collapsing to the single sufficient statistic $\bar{\theta}\alpha$.

C.3 Calibration and Simulation

To quantify the bias, we adopt the Klenow and Willis (2016) Kimball aggregator, the standard parametric MSL-consistent demand system in recent quantitative work:

$$\theta(r_i) \;=\; \bar{\theta}\, r_i^{-\eta}, \qquad \mu(r_i) \;=\; \frac{\theta(r_i)}{\theta(r_i) - 1},$$

where $\eta \geq 0$ is the demand superelasticity in the sense of Klenow and Willis (2016) ($\eta = 0$ recovers the constant-elasticity specification of equation (9)). The implied steady-state pass-through is $\alpha = \eta/(\bar{\theta} - 1)$. We calibrate $(\bar{\theta}, \eta) = (10.86, 0.16)$, corresponding to the aggregate-markup target $M = 1.15$ in Table 1 of Edmond et al. (2023), the central calibration in their quantitative analysis. For the reference value of the transformation elasticity, we use $\sigma = 1.2$, near the median of the IV point estimates in Table 1. Substituting into Proposition 9,

$$\beta \;=\; \frac{1}{\sigma} \cdot \frac{\bar{\theta} - 1}{\bar{\theta} - 1 + \bar{\theta}\eta} \;\approx\; 0.708, \qquad \hat{\sigma} \;\approx\; 1.412,$$

an upward bias in $\hat{\sigma}$ of 17.6%.

We confirm this prediction by Monte Carlo. For each firm f, we draw the main-product demand shock $Z_f \sim U[-0.5, 0]$, the support of which spans the empirical range of main-product buyer lockdown exposure documented in Section 3.3, and small idiosyncratic productivity shocks $\Delta \log a_{m,f}$, $\Delta \log a_{g,f} \sim N(0, 0.05^2)$ independently. We solve the partial equilibrium nonlinearly. Each replication contains 500 firms; we run 2,000 replications per design. Results are reported in Table A6.

Table A6: Monte Carlo Results

Data-generating process	$\hat{\beta}$	$\hat{\sigma}$
Baseline (CET + constant-elasticity demand + constant markup)	0.878	1.202
	(0.226)	(0.265)
Klenow and Willis (2016) Kimball + variable markup	0.733	1.414
	(0.148)	(0.262)

Notes: Means across 2,000 replications, each with 500 firms. Standard deviations across replications in parentheses. Reference value $\sigma = 1.2$ ($1/\sigma = 0.833$).

In the baseline, the estimator recovers σ essentially without bias, confirming the validity of the simulation environment. Under the Klenow–Willis calibration, $\hat{\sigma}$ exhibits the upward bias predicted by Proposition 9, with a simulated magnitude that closely tracks the first-order analytical prediction of 17.6%. The close agreement indicates that nonlinearities induced by the shock size are quantitatively negligible at this calibration.

C.3.1 Implication for the Empirical Estimates

Within the broad class of MSL-consistent variable-markup deviations from constant relative markups (Matsuyama, 2025), Proposition 9 establishes that $\hat{\sigma}$ overstates rather than understates the true transformation elasticity, and the Klenow–Willis calibration shows that the magnitude of this bias is moderate. The IV estimates in Table 1 are therefore conservative for the paper's claim that σ is finite: correcting for this markup-response channel would imply a smaller σ and reinforce the rejection of product-line separability.

D Methodology for Product-Level Markup Estimation

We estimate product-level markups using the production approach of Dhyne et al. (2022), while extending their framework to incorporate a CET production function based on general functional form proposed by Cairncross and Morrow (2023). Dhyne et al. (2022) proposed a production function estimation method that is like that of Ackerberg et al. (2015) yet is based on the production set of Diewert (1973).

D.1 Production Technology Framework

Using the previously estimated elasticity of transformation parameter σ, we specify the CET production function. The firm's technology is characterized by the relationship:

$$F_i^Q\left(\left\{q_{ig}\right\}_{g\in G}\right) = A_i F_i^I\left(M_i, L_i, K_i\right),$$

where F_i^I represents the input aggregation function.

D.2 Cost Minimization and Markup Identification

Following Dhyne et al. (2022), we rely on cost minimization to identify unobserved marginal costs. Firms have inputs $X_i = (M_i, K_i, L_i)$, where M, L are variable inputs and one fixed input, capital K. The cost minimization problem that a firm faces is:

$$C_i(\left\{q_{ig}\right\}_{g\in G}, X_i, P_i) = \min \sum P_i^X X_i \quad \text{s.t.} F_i^Q = F_i^I.$$

The first-order condition for a static input M_s yields:

$$P_i^X + \lambda_i \frac{\partial F_i^I}{\partial M_i} = 0, \tag{18}$$

where λ_i is the relevant Lagrangian multiplier. Marginal costs $MC_{ig} \equiv \frac{\partial C_i}{\partial q_{ig}}$ can be written as:

$$MC_{ig} = -\lambda_i \frac{\partial F_i^Q}{\partial q_{ig}}. \tag{19}$$

Combining equations (18) and (19) yields:

$$\frac{1}{MC_{ig}} = \frac{\frac{\partial F_i^I}{\partial M_i}}{\frac{\partial F_i^Q}{\partial q_{ig}}} \frac{1}{P_i^M}.$$

The firm-product markup $\mu_{ig} \equiv \frac{P_{ig}}{MC_{ig}}$ is obtained by:

$$\mu_{ig} = \frac{\frac{\partial \ln F_i^I(X_i)}{\partial \ln M_i}}{\frac{\partial \ln F_i^Q}{\partial \ln q_{ig}}} \frac{R_{ig}}{P_i^M M_i}.$$

where R_{ig} represents the revenue from product g and $P_i^M M_i$ represents the expenditure on materials. Finally, in the CET case, we specify:

$$F_i^I \equiv Q = \left(\sum_{i=1}^{N} \left(\frac{q_{ig}}{a_{ig}} \right)^{\frac{\sigma+1}{\sigma}} \right)^{\frac{\sigma}{\sigma+1}}$$

Following Dhyne et al. (2022); Cairncross and Morrow (2023) to gain markup identification, we impose the restriction that all products share the same productivity parameter: $a_{ig} = a_i$ for all g. Then the markup of product g becomes:

$$\mu_{ig} = \frac{\frac{\partial \ln F_i^I}{\partial \ln M_i}}{\left(q_{ig}\right)^{\frac{\sigma+1}{\sigma}} / \sum_{g'=1}^{N} \left(q_{ig'}\right)^{\frac{\sigma+1}{\sigma}}} \frac{R_{ig}}{P_i^M M_i}$$

D.3 Estimation

To identify $\frac{\partial \ln F_i^I}{\partial \ln M_i}$, Dhyne et al. (2022) assume that firms use a Cobb-Douglas production function with three factors (Capital K, Labor L, and Materials M). To account for unobserved productivity, we follow Ackerberg et al. (2015)'s control function approach, and we use lagged values as instruments. We estimate the following production function:

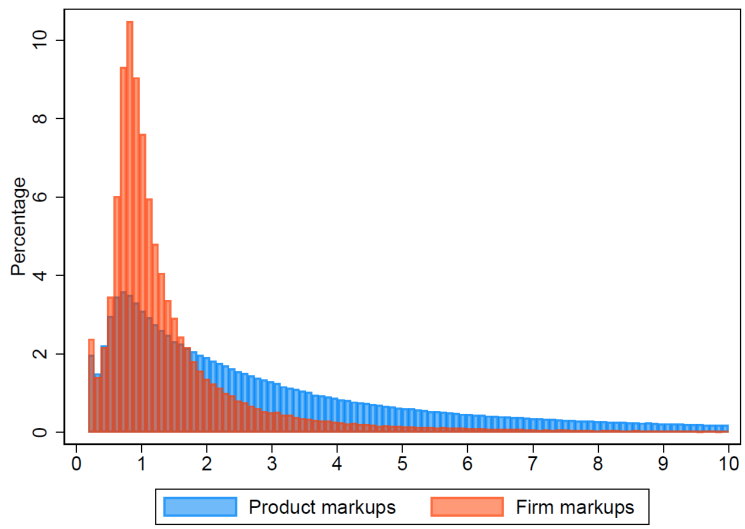
$$\log Q_{it} = \beta_0 + \beta_k \log K_{it} + \beta_l \log L_{it} + \beta_m \log M_{it} + \omega_{it}$$

where ω_{it} represents firm-specific productivity.

To recover quantity measures free of price variation, we construct firm-level price indexes using Tornqvist indices. This enables the estimation of quantity-based rather than revenue-based output elasticities, addressing a key critique in the literature (Bond et al. (2021b)).

We estimate production functions separately for each one-digit product category using GMM with lagged inputs as instruments to recover output elasticities. We leverage the material input elasticity to compute firm-product level markups. Figure A2 presents the markup distribution in 2018 for product versus firm level markups.

Figure A2: Markup estimation results



Notes: The product markup distribution is winsorized at the 1 and 99 percentiles. The firm-level markup is built using product-level cost shares and weights (product value over product-level markup). The graph shows markups until a value of 10, which includes 98% of the firm level markup and 92% of the product level markup.

E Accounting Markups in the Multiproduct Term

As a diagnostic exercise, we assess whether the quantitative importance of the multiproduct term is driven by the product-level markup estimator used to construct own wedges. In the baseline implementation, cumulative wedges are given by

$$\Gamma_{ig} = \frac{\tilde{\lambda}_{ig}}{\lambda_{ig}} \mu_{ig},$$

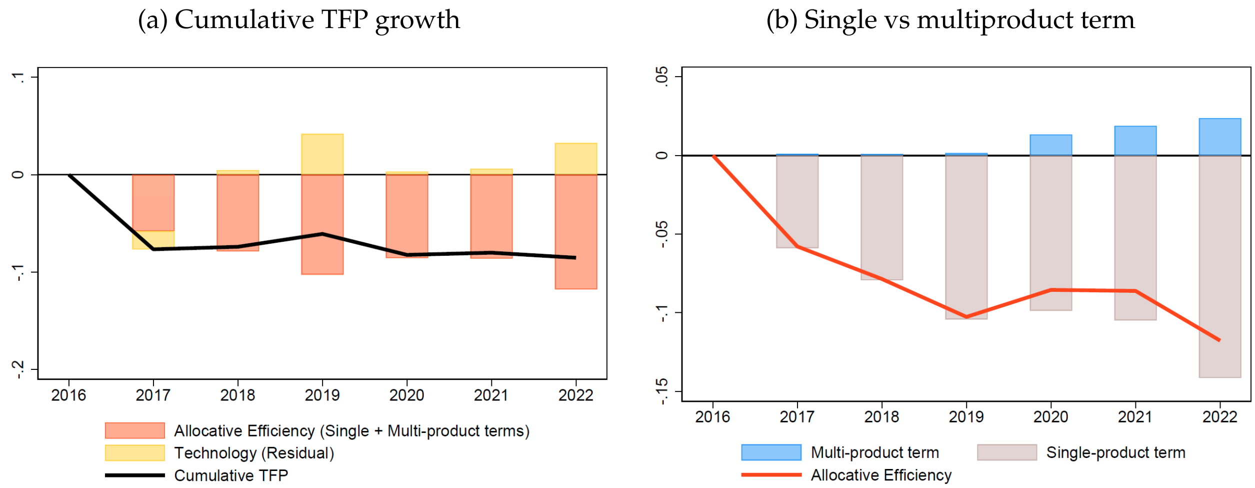
where $\tilde{\lambda}_{ig}/\lambda_{ig}$ captures downstream cumulative wedges and μ_{ig} is the product-level own markup. In this exercise, we leave the single-product component unchanged and modify only the construction of the multiproduct term. Specifically, the factor-share and firm-level markup terms are computed exactly as in the baseline decomposition, using the estimated markups. We then recompute the multiproduct covariance term replacing the product-level own markup μ_{ig} with a firm-level accounting markup, μ_i^{acct}:

$$\Gamma_{ig}^{acct} = \frac{\tilde{\lambda}_{ig}}{\lambda_{ig}} \mu_i^{acct}.$$

This exercise is intentionally conservative. Because μ_i^{acct} assigns the same own markup to all products within a firm, it mechanically compresses within-firm own-wedge heterogeneity. Therefore, the exercise does not provide an alternative product-level markup estimator; instead, it asks how much of the joint-production correction remains when the own-wedge component is measured using a coarser accounting object. Panel A reproduces the main TFP decomposition using this accounting-markup version of the multiproduct term. Incorporating joint production reduces the contribution of allocative efficiency to the TFP decline from 166% under the single-product benchmark to approximately 138% under joint production. In levels, the allocative-efficiency decline falls from about 14% to 11.77%. Thus, even under this coarse own-wedge measure, ignoring joint production overstates the contribution of allocative efficiency to the TFP decline by approximately 19%.

Panel B decomposes the allocative-efficiency component into the single-product term and the multiproduct term. Since the single-product term is held fixed relative to the baseline decomposition, the difference between this exercise and the baseline comes entirely from the alternative construction of cumulative wedges in the multiproduct covariance term. The accounting-markup diagnostic preserves more than half of the baseline joint-production correction, suggesting that the main quantitative conclusion is not driven solely by the product-level markup estimator.

Figure A3: TFP growth decomposition using accounting markups



Notes: Panel (a) shows the cumulative change calculated by applying Corollary 1 repeatedly each year using accounting markups, Panel (b) decomposes the cumulative change in allocative efficiency single-product and multiproduct terms.

Why are the aggregate results so insensitive to how own markups are measured? The reason is that own markups account for only a small share of the variation in cumulative wedges. To see this, we decompose the variance of $\log \Gamma_{ig}$ into the variance of downstream wedges, the variance of own markups, and their covariance. Table A7 reports the decomposition by year: most of the variation comes from downstream distortions, while own markups play a relatively small role. Replacing the product-level own markup with a coarser firm-level object therefore leaves cumulative wedges, and hence the multiproduct term, largely intact.

This pattern is not mechanical. Own markups reflect only the local distortion charged by a single firm-product pair, whereas downstream wedges cumulate distortions along the entire supply chain. Were chains short, or were markups confined to a few isolated stages, downstream wedges would add little dispersion and the two components would be comparable. That they instead dominate the variation in $\log \Gamma_{ig}$ indicates that distortions compound across many successive stages: successive marginalization is quantitatively important in product-firm-level production networks.

Table A7: Variance decomposition of $\log \Gamma$

Year	Downstream wedge	Own markup	Covariance
2016	92.4%	7.3%	0.3%
2017	92.4%	7.4%	0.2%
2018	92.5%	6.9%	0.6%
2019	92.9%	6.6%	0.5%
2020	92.9%	6.2%	0.9%
2021	93.2%	5.4%	1.4%
2022	93.9%	4.9%	1.2%

Notes: We compute the variance decomposition of $\Gamma_{ig} \equiv (\tilde{\lambda}_{ig}/\textit{sales share}_{ig}) \times \mu_{ig}$ for each year. The first term is the variance of downstream wedges, the second the variance of own markups, and the last their covariance. We report the share of total variance explained by each.

F Proofs

Proof of Example 1 in Subsection 2.2. This proof uses Proposition 3. First, the multi-product term is zero since that example has no price variation. To compute the endogenous response of the labor share Λ_L to a change in the markup μ_{21}, we begin by expressing Λ_L solely in terms of the cumulative wedges Γ. Recall that the cumulative wedges are defined as:

$$\Gamma_{11} = \mu_{11}\mu_{21}, \quad \Gamma_{12} = \mu_{12},$$

where μ_{ig} denotes the markup of firm i on product g.

The labor share Λ_L can be written as:

$$\Lambda_L = 1 - \tilde{\lambda}_2\left(1 - \frac{1}{\mu_{12}}\right) - \tilde{\lambda}_1\left(1 - \frac{1}{\mu_{21}}\right) - \tilde{\lambda}_1\frac{1}{\mu_{21}}\left(1 - \frac{1}{\mu_{11}}\right).$$

Simplifying, we have:

$$\Lambda_L = \tilde{\lambda}_2\left(\frac{1}{\mu_{12}}\right) + \tilde{\lambda}_1\left(\frac{1}{\mu_{21}\mu_{11}}\right).$$

Using the definitions of the cumulative wedges, this becomes:

$$\Lambda_L = \tilde{\lambda}_2\left(\frac{1}{\Gamma_{12}}\right) + \tilde{\lambda}_1\left(\frac{1}{\Gamma_{11}}\right).$$

Since Γ_{12} does not depend on μ_{21}, the dependence of Λ_L on μ_{21} is solely through Γ_{11}. Differentiating Λ_L with respect to μ_{21}, we obtain:

$$\frac{d\Lambda_L}{d\mu_{21}} = -\tilde{\lambda}_1 \frac{1}{\Gamma_{11}^2} \mu_{11}.$$

Therefore, the derivative of $\log \Lambda_L$ with respect to $\log \mu_{21}$ is:

$$\frac{d \log \Lambda_L}{d \log \mu_{21}} = \frac{1}{\Lambda_L} \frac{d\Lambda_L}{d\mu_{21}} \mu_{21} = -\frac{\tilde{\lambda}_1 \mu_{11} \mu_{21}}{\Lambda_L \Gamma_{11}^2}.$$

Since $\Gamma_{11} = \mu_{11}\mu_{21}$, we have $\Gamma_{11}^2 = (\mu_{11}\mu_{21})^2$, so the expression simplifies to:

$$\frac{d \log \Lambda_L}{d \log \mu_{21}} = -\frac{\tilde{\lambda}_1}{\Lambda_L \Gamma_{11}}.$$

Substituting the expression for Λ_L in terms of Γ:

$$\Lambda_L = \tilde{\lambda}_2 \left(\frac{1}{\Gamma_{12}}\right) + \tilde{\lambda}_1 \left(\frac{1}{\Gamma_{11}}\right),$$

we find that:

$$\Lambda_L \Gamma_{11} = \tilde{\lambda}_2 \left(\frac{\Gamma_{11}}{\Gamma_{12}}\right) + \tilde{\lambda}_1.$$

Therefore, the derivative simplifies to:

$$\begin{aligned} \frac{d \log \Lambda_L}{d \log \mu_{21}} &= -\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2 \left(\frac{\Gamma_{11}}{\Gamma_{12}}\right) + \tilde{\lambda}_1}, \\ &= -\frac{\tilde{\lambda}_1 \Gamma_{11}^{-1}}{\tilde{\lambda}_2 \Gamma_{12}^{-1} + \tilde{\lambda}_1 \Gamma_{11}^{-1}}, \\ &= -\tilde{\lambda}_1 \frac{\bar{\Gamma}_1}{\Gamma_{11}}. \end{aligned}$$

And

$$\sum_i \tilde{\lambda}_i d \log \mu_i = \lambda_1 d \log \mu_{21}.$$

Therefore,

$$d \log TFP = \tilde{\lambda}_{11} \left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1\right) d \log \mu_{21}.$$

□

Proof of an Example 2 in Subsection 2.2. Let us pick product 2 to be a reference product for firm 1.

Then, we obtain

$$d \log p_{11}/p_{12} = d \log \mu_{11}/\mu_{12} + \frac{1}{\sigma} d \log y_{11}/y_{12}.$$

Using the relation $d \log \lambda = d \log p + d \log y$, we derive

$$d \log p_{11}/p_{12} = \left(\frac{\sigma}{\sigma+1}\right) d \log \mu_{11}/\mu_{12} + \frac{1}{\sigma+1} d \log \lambda_{11}/\lambda_{12},$$

where $d \log \mu_{11} = 0$ and $d \log \mu_{12} = 0$. From the Cobb-Douglas assumption, we know that $d \log(\lambda_{11}/\lambda_{12}) = -d \log \mu_{21}$. Therefore, we have

$$d \log(p_{11}/p_{12}) = -\left(\frac{1}{\sigma+1}\right) d \log \mu_{21}.$$

Now, we can write

$$\begin{aligned}
\mathrm{Cov}_{s_i}\left(d \log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right) &= \mathrm{Cov}_{[\tilde{\lambda}_1, \tilde{\lambda}_2]}\left(\begin{bmatrix} d \log p_{11} \\ d \log p_{12} \end{bmatrix}, \begin{bmatrix} \frac{\Gamma_1}{\Gamma_{11}} \\ \frac{\Gamma_1}{\Gamma_{12}} \end{bmatrix}\right) \\
&= \mathrm{Cov}_{[\tilde{\lambda}_1, \tilde{\lambda}_2]}\left(\begin{bmatrix} d \log p_{11}/p_{12} \\ d \log p_{12}/p_{12} \end{bmatrix}, \begin{bmatrix} \frac{\Gamma_1}{\Gamma_{11}} \\ \frac{\Gamma_1}{\Gamma_{12}} \end{bmatrix}\right) \\
&= \mathrm{Cov}_{[\tilde{\lambda}_1, \tilde{\lambda}_2]}\left(\begin{bmatrix} \left(\frac{1}{\sigma+1}\right) d \log \mu_{21} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\Gamma_1}{\Gamma_{11}} \\ \frac{\Gamma_1}{\Gamma_{12}} \end{bmatrix}\right).
\end{aligned}$$

Observe that

$$\mathbb{E}_{[\tilde{\lambda}_1, \tilde{\lambda}_2]}\begin{pmatrix} \left(\frac{1}{\sigma+1}\right) d \log \mu_{21} \frac{\Gamma_1}{\Gamma_{11}} \\ 0 \end{pmatrix} = \left(\frac{1}{\sigma+1}\right) d \log \mu_{21} \frac{\Gamma_1}{\Gamma_{11}} \tilde{\lambda}_{11},$$

and

$$\begin{aligned}
\mathbb{E}_{[\tilde{\lambda}_1, \tilde{\lambda}_2]}\begin{pmatrix} d \log p_{11}/p_{12} \\ d \log p_{12}/p_{12} \end{pmatrix} \mathbb{E}_{[\tilde{\lambda}_1, \tilde{\lambda}_2]}\begin{pmatrix} \frac{\Gamma_1}{\Gamma_{11}} \\ \frac{\Gamma_1}{\Gamma_{12}} \end{pmatrix} &= \left(\frac{1}{\sigma+1}\right) d \log \mu_{21} \tilde{\lambda}_1 \left(\tilde{\lambda}_1 \frac{\Gamma_1}{\Gamma_{11}} + \tilde{\lambda}_2 \frac{\Gamma_1}{\Gamma_{12}}\right) \\
&= \left(\frac{1}{\sigma+1}\right) d \log \mu_{21} \tilde{\lambda}_1.
\end{aligned}$$

Therefore, the multi-product term is given by

$$\mathrm{Cov}_{s_i}\left(d \log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right) = -\left(\frac{1}{\sigma+1}\right) \tilde{\lambda}_{11} \left(\frac{\bar{\Gamma}_1}{\Gamma_{11}} - 1\right) d \log \mu_{21}.$$

Using the single product term's result,

$$-\underbrace{\sum_{i\in N}\tilde{\lambda}_i d\log\mu_i}_{\text{Firm-level Markup}} - \underbrace{d\log\Lambda_f}_{\text{Aggregate Labor Shares}} = \tilde{\lambda}_{11}\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)d\log\mu_{21}.$$

Finally, we obtain

$$\Delta TFP = \left(1-\frac{1}{\sigma+1}\right)\tilde{\lambda}_{11}\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)d\log\mu_{21}.$$

□

Proof of an Example with Taste Shocks in Subsection 2.2.4: The proof follows similar steps to the markup shock case. Let us first calculate how the labor share responds to taste shocks. Starting from the labor share expression:

$$\Lambda_L = \tilde{\lambda}_2\left(\frac{1}{\Gamma_{12}}\right)+\tilde{\lambda}_1\left(\frac{1}{\Gamma_{11}}\right)=\frac{1}{\bar{\Gamma}_1},$$

differentiating with respect to $\tilde{\lambda}_1$ and using the fact that $d\tilde{\lambda}_2 = -d\tilde{\lambda}_1$, we obtain:

$$d\log\Lambda_L = d\log\tilde{\lambda}_1\left(1-\frac{1}{\tilde{\lambda}_2}+\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}\frac{\bar{\Gamma}_1}{\Gamma_{11}}\right) = d\log\tilde{\lambda}_1\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)$$

Next, to express the multi-product term as a covariance, we write:

$$\begin{aligned}
\mathrm{Cov}_{s_1}\left(d\log p_{(1,\cdot)},\frac{\bar{\Gamma}_1}{\Gamma_{(1,\cdot)}}\right) &= \mathrm{Cov}_{[\tilde{\lambda}_1,\tilde{\lambda}_2]}\left(\begin{bmatrix} d\log p_{11}/p_{12} \\ d\log p_{12}/p_{12}\end{bmatrix},\begin{bmatrix}\frac{\bar{\Gamma}_1}{\Gamma_{11}} \\ \frac{\bar{\Gamma}_1}{\Gamma_{12}}\end{bmatrix}\right) \\
&= \mathrm{Cov}_{[\tilde{\lambda}_1,\tilde{\lambda}_2]}\left(\begin{bmatrix} \frac{1}{\sigma+1}\frac{1}{\tilde{\lambda}_2}d\log\tilde{\lambda}_1 \\ 0\end{bmatrix},\begin{bmatrix}\frac{\bar{\Gamma}_1}{\Gamma_{11}} \\ \frac{\bar{\Gamma}_1}{\Gamma_{12}}\end{bmatrix}\right) \\
&= \frac{1}{\sigma+1}\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}\left(\frac{\bar{\Gamma}_1}{\Gamma_{11}}-1\right)d\log\tilde{\lambda}_1.
\end{aligned}$$

Combining these expressions with equation (5) yields the result. □

Proof of Proposition 3

Lemma 2. *The within-firm cost share equals the cost-based Domar weight share rescaled by cumulative wedges:*

$$c_{ig} = s_{ig}\frac{\bar{\Gamma}_i}{\Gamma_{ig}}, \qquad c_{ig}\equiv\frac{q_{ig}mc_{ig}}{C_i(\boldsymbol{q}_i,\boldsymbol{p})}.$$

Hence $\mathbb{E}_{\boldsymbol{s}_i}[\bar{\Gamma}_i/\Gamma_{(i,\cdot)}] = 1$*, and for any vector* $\boldsymbol{x}_i$*,*

$$\sum_g \left(c_{ig} - s_{ig}\right) x_{ig} = \text{Cov}_{s_i}\left(\boldsymbol{x}_i, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right).$$

Proof. By definition,

$$c_{ig} = \frac{q_{ig} mc_{ig}}{\sum_h q_{ih} mc_{ih}} = \frac{\lambda_{ig}/\mu_{ig}}{\sum_h \lambda_{ih}/\mu_{ih}}.$$

Using $\Gamma_{ig} = \left(\tilde{\lambda}_{ig}/\lambda_{ig}\right)\mu_{ig}$, so that $\lambda_{ig}/\mu_{ig} = \tilde{\lambda}_{ig}/\Gamma_{ig}$,

$$c_{ig} = \frac{\tilde{\lambda}_{ig}/\Gamma_{ig}}{\sum_h \tilde{\lambda}_{ih}/\Gamma_{ih}} = s_{ig}\frac{\bar{\Gamma}_i}{\Gamma_{ig}}.$$

Summing over g gives $\sum_g s_{ig}\left(\bar{\Gamma}_i/\Gamma_{ig}\right) = \sum_g c_{ig} = 1$. □

Lemma 3. *The cost-share-weighted change in marginal cost satisfies*

$$\sum_g c_{ig}\, d\log mc_{ig} = -d\log A_i + \sum_{j,g'} \tilde{\Omega}_{i,jg'} d\log p_{jg'} + \sum_f \tilde{\Omega}_{i,f} d\log w_f.$$

Proof. By definition, $C_i(\boldsymbol{q}_i, \boldsymbol{p}) = \sum_g q_{ig} mc_{ig}$. Differentiating both sides of $\log C_i = \log \sum_g q_{ig} mc_{ig}$ gives, on the one hand,

$$d\log C_i = -d\log A_i + \sum_g c_{ig} d\log q_{ig} + \sum_{j,g'} \tilde{\Omega}_{i,jg'} d\log p_{jg'} + \sum_f \tilde{\Omega}_{i,f} d\log w_f,$$

where Shephard's lemma ($\partial C_i/\partial p_{jg'} = x_{i,jg'}$ and $\partial C_i/\partial w_f = L_{if}$) and the productivity sensitivity ($\partial C_i/\partial A_i = -C_i/A_i$) are used. On the other hand,

$$d\log C_i = \sum_g c_{ig}\left(d\log q_{ig} + d\log mc_{ig}\right).$$

Equating yields the result. □

Lemma 4. *Under constant returns to scale,*

$$\sum_g c_{ig}\, d\log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) = 0.$$

Proof. Since F_i^Q is homogeneous of degree one in $\boldsymbol{q}_i$, its gradient $\partial F_i^Q/\partial q_{ig}$ is homogeneous of degree zero. Euler's theorem applied to $\partial F_i^Q / \partial q_{ih}$ yields

$$\sum_g q_{ig} \frac{\partial^2 F_i^Q}{\partial q_{ig} \partial q_{ih}} = 0 \quad \forall h.$$

Cost minimization implies $mc_{ig} = \xi_i \, \partial F_i^Q / \partial q_{ig}$ for some firm-level multiplier ξ_i, and Euler's theorem applied to F_i^Q gives $C_i = \sum_g q_{ig} mc_{ig} = \xi_i F_i^Q$. Hence

$$c_{ig} = \frac{q_{ig} \, \partial F_i^Q / \partial q_{ig}}{F_i^Q}.$$

Differentiating,

$$d \log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) = \sum_h \frac{q_{ih}}{\partial F_i^Q / \partial q_{ig}} \frac{\partial^2 F_i^Q}{\partial q_{ig} \partial q_{ih}} d \log q_{ih},$$

and therefore

$$\sum_g c_{ig} d \log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) = \frac{1}{F_i^Q} \sum_h q_{ih} d \log q_{ih} \sum_g q_{ig} \frac{\partial^2 F_i^Q}{\partial q_{ig} \partial q_{ih}} = 0.$$

□

Proof of Proposition 3. By Hall (1973), cost minimization implies $mc_{ig}/mc_{ih} = \left(\partial F_i^Q / \partial q_{ig}\right) / \left(\partial F_i^Q / \partial q_{ih}\right)$. Taking log differences, averaging over h with weights c_{ih}, and applying Lemma 4,

$$d \log mc_{ig} - \sum_h c_{ih} d \log mc_{ih} = d \log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right).$$

Combining with Lemma 3 and using $d \log p_{ig} = d \log \mu_{ig} + d \log mc_{ig}$ gives the product-level price equation:

$$d \log p_{ig} = -d \log A_i + d \log \mu_{ig} + d \log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) + \sum_{j,g'} \tilde{\Omega}_{i,jg'} d \log p_{jg'} + \sum_f \tilde{\Omega}_{i,f} d \log w_f.$$

Define $d \log \mu_i \equiv \sum_g c_{ig} d \log \mu_{ig}$, the first-order expansion of $\mu_i = \text{sales}_i / \text{total cost}_i$ around the efficient equilibrium. Cost-share averaging the product-level price equation and applying Lemma 4,

$$\sum_g c_{ig} d \log p_{ig} = -d \log A_i + d \log \mu_i + \sum_{j,g'} \tilde{\Omega}_{i,jg'} d \log p_{jg'} + \sum_f \tilde{\Omega}_{i,f} d \log w_f.$$

Subtracting the firm-level equation from the product-level equation yields the within-firm deviation

$$d\log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) = \left(d\log p_{ig} - \sum_h c_{ih} d\log p_{ih}\right) - \left(d\log\mu_{ig} - d\log\mu_i\right).$$

Stacking the product-level price equation over (i,g) and using $d\log Y = -\boldsymbol{b}' d\log\boldsymbol{p}$ together with $\tilde{\boldsymbol{\lambda}}' = \boldsymbol{b}'(I-\tilde{\Omega})^{-1}$,

$$d\log Y = \sum_i \tilde{\lambda}_i d\log A_i - \sum_{ig}\tilde{\lambda}_{ig} d\log\mu_{ig} - \sum_{ig}\tilde{\lambda}_{ig} d\log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) - \sum_f \tilde{\Lambda}_f d\log w_f.$$

Substituting the within-firm deviation and using $s_{ig} = \tilde{\lambda}_{ig}/\tilde{\lambda}_i$,

$$\sum_{ig}\tilde{\lambda}_{ig} d\log\left(\frac{\partial F_i^Q}{\partial q_{ig}}\right) = \sum_i \tilde{\lambda}_i \sum_g \left(s_{ig} - c_{ig}\right) d\log p_{ig} - \sum_{ig}\tilde{\lambda}_{ig} d\log\mu_{ig} + \sum_i \tilde{\lambda}_i d\log\mu_i.$$

By Lemma 2,

$$\sum_i \tilde{\lambda}_i \sum_g \left(s_{ig} - c_{ig}\right) d\log p_{ig} = -\sum_i \tilde{\lambda}_i \operatorname{Cov}_{\boldsymbol{s}_i}\left(d\log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right).$$

Substituting and subtracting the factor-quantity terms gives

$$d\log TFP = \sum_i \tilde{\lambda}_i \operatorname{Cov}_{\boldsymbol{s}_i}\left(d\log p_{(i,\cdot)}, \frac{\bar{\Gamma}_i}{\Gamma_{(i,\cdot)}}\right) - \sum_f \tilde{\Lambda}_f d\log\Lambda_f - \sum_i \tilde{\lambda}_i d\log\mu_i + \sum_i \tilde{\lambda}_i d\log A_i.$$

□

Proof of Proposition 4. Let $i = 1,\dots,N$ index N products. Each product i faces the isoelastic demand

$$q_i \;=\; D_i p_i^{-\theta_i}, \quad \theta_i > 1,$$

and its price satisfies $p_i = \mu_i \frac{\partial C}{\partial q_i}$, where $\mu_i > 0$ is a wedge (e.g. a markup). The cost function

$$C(q_1,\dots,q_N) \;=\; \frac{w}{A}\Big(\sum_{j=1}^N q_j^{\frac{\sigma+1}{\sigma}}\Big)^{\frac{\sigma}{\sigma+1}}, \quad \sigma > 0,$$

implies that

$$d\log p_i \;=\; d\log\Big[\mu_i \tfrac{\partial C}{\partial q_i}\Big] \;=\; d\log\Big[\tfrac{\partial C}{\partial q_i}\Big] \quad (\mu_i \text{ fixed}).$$

To derive this marginal-cost term explicitly, define

$$S_j \;:=\; \frac{q_j^{(\sigma+1)/\sigma}}{\sum_{\ell=1}^N q_\ell^{(\sigma+1)/\sigma}}, \quad \sum_{j=1}^N S_j \;=\; 1,$$

and let

$$\overline{d\log q} \ := \ \sum_{j=1}^{N} S_j \, d\log q_j.$$

A standard differentiation of the CET cost function shows that

$$d\log p_i \ = \ \frac{1}{\sigma}\Big[d\log q_i \ - \ \overline{d\log q}\Big]. \tag{20}$$

Since the shock is exclusively to product k, we have $d\log D_k < 0$ and $d\log D_i = 0$ for $i \neq k$. From the isoelastic demand,

$$d\log q_i \ + \ \theta_i \, d\log p_i = d\log D_i,$$

so for $i \neq k$,

$$d\log q_i + \theta_i \, d\log p_i \ = \ 0, \quad \text{and for } i = k, \quad d\log q_k + \theta_k \, d\log p_k \ = \ d\log D_k < 0.$$

Combining $d\log q_i + \theta_i \, d\log p_i = 0$ with (20) gives

$$\theta_i\Big[\tfrac{1}{\sigma}\Big(d\log q_i - \overline{d\log q}\Big)\Big] \ = \ -\, d\log q_i \quad (i \neq k),$$

which simplifies to

$$(\theta_i + \sigma)\, d\log q_i \ = \ \theta_i \, \overline{d\log q} \quad (i \neq k).$$

An analogous expression arises for $i = k$, except that the right-hand side involves $d\log D_k$:

$$(\theta_k + \sigma)\, d\log q_k \ = \ \sigma \, d\log D_k \ + \ \theta_k \, \overline{d\log q}.$$

Solving these equations jointly forces $\overline{d\log q}$ to have the same sign as $d\log D_k$. In particular, since $d\log D_k < 0$ by assumption, one can verify that the unique solution consistent with marginal-cost equality implies

$$\overline{d\log q} \ < \ 0.$$

Then for each $i \neq k$, the equation

$$(\theta_i + \sigma)\, d\log q_i \ = \ \theta_i \, \overline{d\log q}$$

implies $d\log q_i < 0$ (because both $\theta_i + \sigma > 0$ and $\theta_i > 0$). This confirms part (i) of the proposition.

Finally, we substitute $d\log q_i < 0$ and $\overline{d\log q} < 0$ into (20) to find

$$d\log p_i \ = \ \frac{1}{\sigma}\Big[d\log q_i - \overline{d\log q}\Big].$$

Since $|d\log q_i| \ < \ |\overline{d\log q}|$ for $i \neq k$ but $d\log q_i$ and $\overline{d\log q}$ are both negative, the difference in brackets is strictly positive, and hence $d \log p_i > 0$. This establishes part (ii).

Thus for each product $g \neq k$, we have $d \log q_g < 0$ and $d \log p_g > 0$ when $d \log D_k < 0$. ◻

Proof of Proposition 7. From the resource constraint,

$$q_{ig} = y_{ig} + \sum_{j \in N} x_{jig}.$$

$$d \log y_{ig} = \frac{q_{ig}}{y_{ig}} d \log q_{ig} - \sum_j \frac{x_{jig}}{y_{ig}} d \log x_{jig}.$$

From the cost minimization of joint production we know

$$\sum_g q_{ig} mc_{ig} d \log q_{ig} = \sum_{jp} x_{i,jp} p_{jp} d \log x_{i,jp} + \sum_f w_f x_{i,f} d \log L_{if}.$$

$$\sum_{jp} \frac{x_{i,jp} p_{jp}}{GDP} d \log x_{i,jp} = \sum_g \frac{1}{\mu_{ig}} \frac{q_{ig} p_{ig}}{GDP} d \log q_{ig} - \sum_f \frac{w_f x_{i,f}}{GDP} d \log L_{if}.$$

By cost minimization assumption,

$$\begin{aligned}
d \log Y &= \sum_{ig} \frac{p_{ig} y_{ig}}{GDP} d \log y_{ig} - \sum_f \frac{w_f x_{i,f}}{q_{ir} mc_{ig}} d \log L_f, \\
&= \sum_{ig} \frac{p_{ig} y_{ig}}{GDP} \left(\frac{q_{ir}}{y_{ig}} d \log q_{ig} - \sum_j \frac{x_{jig}}{y_{ig}} d \log x_{jig} \right) - \sum_f \frac{w_f L_f}{GDP} d \log L_f, \\
&= \sum_{ig} \frac{p_{ig} y_{ig}}{GDP} \frac{q_{ir}}{y_{ig}} d \log q_{ig} - \sum_{ig} \sum_j \frac{p_{ig} y_{ig}}{GDP} \frac{x_{jig}}{y_{ig}} d \log x_{jig} - \sum_f \frac{w_f L_f}{GDP} d \log L_f, \\
&= \sum_{ig} \frac{p_{ig} q_{ig}}{GDP} d \log q_{ig} - \sum_{ig} \frac{1}{\mu_{ig}} \frac{p_{ig} q_{ig}}{GDP} d \log q_{ig} + \sum_{if} \frac{w_f x_{i,f}}{GDP} d \log L_{if} - \sum_f \frac{w_f L_f}{GDP} d \log L_f, \\
&= \sum_{ig} \lambda_{ig} \left(1 - \mu_{ig}^{-1}\right) d \log q_{ig}.
\end{aligned}$$

At the efficient equilibrium, $\mu_{ig} = 1$, so $1 - \mu_{ig}^{-1} = 0$ and the first-order effect on TFP from introducing markup wedges vanishes; the leading effect is second order. In the expression above, terms involving changes in λ_{ig} or higher-order quantity responses are multiplied by $1 - \mu_{ig}^{-1}$ and therefore vanish at the efficient point. The only surviving first-order variation of the wedge term is

$$d\left(1 - \mu_{ig}^{-1}\right) = \mu_{ig}^{-1} d \log \mu_{ig} = d \log \mu_{ig} \quad \text{at } \mu_{ig} = 1.$$

The second differential of log TFP is therefore

$$d^2 \log TFP = \sum_{ig} \lambda_{ig} \, d \log q_{ig} \, d \log \mu_{ig}.$$

Since the first-order term vanishes, a second-order Taylor expansion gives

$$\Delta \log TFP = \frac{1}{2} \sum_{ig} \lambda_{ig} \, d \log q_{ig} \, d \log \mu_{ig}.$$

Since the welfare loss is $\mathcal{L} = -\Delta \log TFP$,

$$\mathcal{L} = -\frac{1}{2} \sum_{ig} \lambda_{ig} \, d \log q_{ig} \, d \log \mu_{ig}.$$

□

Proof of Proposition 8. Let

$$m_{ig} \equiv d \log \mu_{ig}, \qquad m_i \equiv \sum_g s_{ig} m_{ig}, \qquad \bar{m} \equiv \sum_i \lambda_i m_i = \sum_i \sum_g \lambda_{ig} m_{ig},$$

where

$$\lambda_i = \sum_g \lambda_{ig}, \qquad s_{ig} = \frac{\lambda_{ig}}{\lambda_i}.$$

Let

$$\hat{p}_{ig} \equiv d \log p_{ig}, \qquad \hat{q}_{ig} \equiv d \log q_{ig}, \qquad \hat{w} \equiv d \log w.$$

We normalize the aggregate CES price index so that $d \log P = 0$.

Because the expansion is around an efficient equilibrium, product-level revenue shares and product-level cost shares coincide. Hence s_{ig} is also the local cost share of product g within firm i. Define

$$\hat{p}_i \equiv \sum_g s_{ig} \hat{p}_{ig}, \qquad \hat{q}_i \equiv \sum_g s_{ig} \hat{q}_{ig}.$$

For a firm with CET output technology, the local marginal-cost response is

$$d \log mc_{ig} = \hat{w} + \frac{1}{\sigma} \left(\hat{q}_{ig} - \hat{q}_i \right).$$

Since $\hat{p}_{ig} = d \log mc_{ig} + m_{ig}$, we have

$$\hat{p}_{ig} = \hat{w} + m_{ig} + \frac{1}{\sigma} \left(\hat{q}_{ig} - \hat{q}_i \right).$$

Averaging within firm i using weights s_i gives

$$\hat{p}_i = \hat{w} + m_i.$$

Subtracting this firm-level average from the product-level price equation yields

$$\hat{p}_{ig} - \hat{p}_i = m_{ig} - m_i + \frac{1}{\sigma}\left(\hat{q}_{ig} - \hat{q}_i\right).$$

CES final demand implies the within-firm relative demand equation

$$\hat{q}_{ig} - \hat{q}_i = -\theta\left(\hat{p}_{ig} - \hat{p}_i\right).$$

Substituting this condition into the previous equation gives

$$\hat{p}_{ig} - \hat{p}_i = m_{ig} - m_i - \frac{\theta}{\sigma}\left(\hat{p}_{ig} - \hat{p}_i\right).$$

Therefore,

$$\hat{p}_{ig} - \hat{p}_i = \frac{\sigma}{\sigma + \theta}\left(m_{ig} - m_i\right).$$

Define

$$\kappa \equiv \frac{\sigma}{\sigma + \theta}.$$

Then the product-level price response can be written as

$$\hat{p}_{ig} = \hat{w} + m_i + \kappa(m_{ig} - m_i).$$

Using the aggregate CES price-index normalization,

$$0 = d\log P = \sum_i \sum_g \lambda_{ig}\hat{p}_{ig}.$$

Substituting the price response gives

$$0 = \hat{w} + \sum_i \lambda_i m_i + \kappa \sum_i \sum_g \lambda_{ig}(m_{ig} - m_i).$$

The last term is zero by the definition of m_i. Hence

$$\hat{w} = -\bar{m}.$$

Therefore,

$$\hat{p}_{ig} = m_i - \bar{m} + \kappa(m_{ig} - m_i).$$

Since $d \log P = 0$, CES demand gives

$$\hat{q}_{ig} = -\theta \hat{p}_{ig}.$$

Using Proposition 7,

$$\mathcal{L} = -\frac{1}{2} \sum_i \sum_g \lambda_{ig} \hat{q}_{ig} m_{ig}.$$

Thus

$$\mathcal{L} = \frac{1}{2} \theta \sum_i \sum_g \lambda_{ig} \hat{p}_{ig} m_{ig}.$$

Substituting the expression for $\hat{p}_{ig}$,

$$\mathcal{L} = \frac{1}{2} \theta \sum_i \sum_g \lambda_{ig} \left[m_i - \bar{m} + \kappa (m_{ig} - m_i) \right] m_{ig}.$$

The first component is

$$\sum_i \sum_g \lambda_{ig} (m_i - \bar{m}) m_{ig} = \sum_i \lambda_i (m_i - \bar{m}) m_i = \mathrm{Var}_{\bar{\lambda}}(m_i).$$

The second component is

$$\sum_i \sum_g \lambda_{ig} (m_{ig} - m_i) m_{ig} = \sum_i \lambda_i \, \mathrm{Var}_{s_i}(m_{ig}).$$

Therefore,

$$\mathcal{L} = \frac{1}{2} \theta \left[\mathrm{Var}_{\bar{\lambda}}(m_i) + \frac{\sigma}{\sigma + \theta} \mathbb{E}_{\bar{\lambda}} \left\{ \mathrm{Var}_{s_i}(m_{ig}) \right\} \right].$$

Finally, by the law of total variance,

$$\mathrm{Var}_{\lambda}(m_{ig}) = \mathrm{Var}_{\bar{\lambda}}(m_i) + \mathbb{E}_{\bar{\lambda}} \left\{ \mathrm{Var}_{s_i}(m_{ig}) \right\}.$$

Hence,

$$\mathcal{L} = \frac{1}{2} \theta \left[\mathrm{Var}_{\lambda}(m_{ig}) - \frac{\theta}{\sigma + \theta} \mathbb{E}_{\bar{\lambda}} \left\{ \mathrm{Var}_{s_i}(m_{ig}) \right\} \right].$$

□